

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:29:25 ; Search time 1907.71 Seconds  
(without alignments)  
12741.429 Million cell updates/sec

Title: US-09-749-589-1

Perfect score: 2262  
Sequence: 1 atgagccagcagccagccgcg.....agacctgacgcgcctgtga 2262

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estol:\*  
6: em\_estro:\*  
7: em\_estov:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_estl:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rnd:\*  
20: em\_gss\_vtl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	269.6	11.9	566	10	BE168072 QV3-HT051
2	258.6	11.4	343	11	BG385110 306720 MA
3	189	8.4	837	11	BG743866 602722621
C 4	188.6	8.3	386	11	BF358998 CM1-ET004
5	179.2	7.9	1246	12	AK018566 Mus muscu
C 6	170.8	7.6	622	13	AZ833796 2M0116E10
7	158.6	7.0	586	10	BE031984 130773 MA
8	135.4	6.0	857	11	BF578820 602094862
9	132.8	5.9	611	11	BG680597 602629036
10	129.2	5.7	425	10	BE156730 QV0-HT036
11	125.8	5.6	283	10	BB605137 BB605137
12	123.6	5.5	521	11	BG732794 346420 MA

13	122.4	5.4	481	11	BF463964
14	121.8	5.4	558	11	BG727939
15	120.8	5.3	510	10	AA871419 VQ35D04.T
16	119.2	5.3	465	11	BF890309
17	117.8	5.2	524	10	AM656395 108633 MA
18	116	5.1	237	10	BB570880
19	115	5.1	218	11	BF883081
20	111	4.9	595	10	BE156600
C 21	106	4.7	469	13	AI006678
C 22	106	4.7	478	13	AQ976378
23	106	4.7	558	10	AI037299
24	105.8	4.7	487	11	BF564736
25	104	4.6	321	10	AM436999
26	103.6	4.6	392	10	AM480438
C 27	101.6	4.5	385	10	AM142905
C 28	100	4.4	688	11	BG863674
C 29	98	4.3	682	11	BF610846
30	97.4	4.3	496	11	BF441578
31	97.4	4.3	497	11	BF441641
32	96.8	4.3	488	10	AA871638
C 33	93.8	4.1	521	13	AZ066603
C 34	93.4	4.1	1110	11	BF983692
35	92.8	4.1	568	13	AZ966598
36	89.6	4.0	310	10	AA717241
37	89.4	4.0	716	11	BG718763
38	87.4	3.9	589	10	AM611310
39	87.4	3.9	661	11	BI220929
40	87.4	3.9	784	11	BI330869
41	87.4	3.9	915	11	BI219321
42	87	3.8	556	10	AM603701
43	86.4	3.8	515	11	BG815100
44	85.8	3.8	653	11	BI217880
45	85.6	3.8	799	11	BG966791

#### ALIGNMENTS

RESULT 1  
LOCUS BE168072 569 bp mRNA EST 21-JUN-2000  
DEFINITION QV3-HT0513-150300-117-f10 HT0513 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BE168072  
VERSION BE168072.1 GI:8630793  
KEYWORDS  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 569)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W.J., Zago,M.A., Bordin,S., Costa,F.P.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.O. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE JOURNAL  
COMMENT MEDLINE  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-HT0513-150  
300-117-f10&t3=2000-03-15&t4=1)  
Seq primer: puc 18 forward

BASE COUNT	74 a	112 c	81 g	76 t
ORIGIN				

Query Match	11.48;	Score 258.6;	DB 11;	length 343;
Best Local Similarity	85.68;	Pred. NO. 4.1e-46;		
Matches 304;	Conservative 0;	Mismatches 39;	Indels 12;	Gaps 1;
1662	tgctccctgagagagctgcgcagcagactllgagatgagcccccacagacccaacaa	1721		

Dy 1662 tgtctccctcaggggcttgacacgaactttgagaattgcgcccccacgacccccaca 1721

Db 1 TGTCTCCCTCAGGAGACTGCACGACGACTTTGAGATGTCCTTCCCACCGACCCCAACA 60

QY 1722 caaccagaccccgctaacgycaccagcgtgtcctatatcaccttcagccctgacagctc 1781

Db 61 CAACCAGACGCCCTGCCAATGGCCGAGCGTGTCCTACATCACCCTCAGCCCTGACAGCTC 120

QY 1782 ctcaactgcccagagtgaaccacagcctccgctgagcccccggcagagcccaagtgcacat 1841

Db 121 CACAGCTGTCCCTTGTGAGCCCCCTGGCCTCCGCT-----GAGCCGAGGGACAT 168

QY 1842 gctggccagcgctccacccttcgtcaccttccacacccctcatcctggaacatgagtggagt 1901

Db 169 GCTGGCCAGCGTCCCAACCCCTTTGTCACTTCCACACCCCTCATCCCTGGACATGAGTGGCAT 228

QY 1902 cagcttcgtgacttgatggcatcaagccctggccaagctgagctccaccctatgggaa 1961

db 229 CAGCTTGTGGACCTGATGGGAATCAAGCTCTGGATAAGCTGAGCTCCACTTACAGGAA 288

1962 gatcgcgc tgaagg tctctct tgg tgaacatccatgc ccaggtg tacaatgacatc 2016 •

DB 289 GATTGGTGGTGCAGGTCCTTTTGGTGAACATTCATGGCCAGGTGTACAAAGACATTT 343

### RESULT 3

RG743866	LOCUS	837 bp	mRNA	EST	15-MAY-2001
----------	-------	--------	------	-----	-------------

DEFINITION	602/22621.F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849228 5',
MRNA sequence	

ACCESSION	BG743866
VERSION	BG743866.1
	GT:14054519

**KEYWORDS** EST. human SOURCE

ORGANISM	Homo sapiens
Eukaryota:	Metazoa:
Chordata:	Vertebrata:
Euteleostomi:	

REFERENCE  
1 (bases 1 to 837)  
Mammalia; Eulheria; Primates; Catarrhini; Hominoidea; Homo.

**AUTHORS** NIH-MGC <http://mgc.ncl.nih.gov/>;  
**TITLE** National Institutes of Health: Mammalian Gene Collection (MGC)

COMMENT  
JOURNAL  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-1@mail.hih.gov](mailto:cgapbs-1@mail.hih.gov)  
Tissue Procurement: Dr. Daniel McVicar, PBS/NCT

CDNA Library Prepared by: Ling Hong/Kublin Laboratory  
CDNA Library Arraved by: The I M A G E Consortium (I.M.I.)

Clone distribution: MGC clone distribution information can be  
DNA sequencing by: Incyte Genomics, Inc.

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLCML688 Row: 1 Column: 05  
High quality sequence store: 807

## FEATURES

### Location/Qualifiers

	source	1..837 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4849228" /clone_id="NIH_MGC_106" /tissue_type="natural killer cells, cell line" /lab_host="DH10B (phage-resistant)" Note:"Organ: blood; Vector: pORF9; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and BspScrip II RT (Life Technologies). Note: this is a NIH_MGC library."
BASE COUNT	142 a      249 c      254 g      192 t	
ORIGIN		
Query Match	8.4%; Score 189; DB 11; Length 837;	
Best Local Similarity	53.5%; Pred. No. 6.8e-31.	
Matches 416; Conservative	0; Mismatches 360; Indels 1; Gaps 1	
OY 714	ccttaacctcatgtacattgcgcaaaaccctcccccaaccaaacatcgctcgtcatactt 773	
Db 2	CTATACATTGTCTGGAGGTCGTCTGGAACTGCCCCAGAACAGTGTTGGCACCGTGTCA 61	
OY 774	cgcctcacgaugtgcttcctcgttgtcgtgtgtagaaggagccatagtctcgctaatca 833	
Db 62	TGCACACTGTGCTGGGGGTGTGCTGTGTGTGAAGCTGTTGAATGACAACTGCACGA 121	
OY 834	caagaatcgcttcccactccctacaagatatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 893	
Db 122	GCAAGTGGCCAATGCCAATACCAGGAGGAGCTGCTACGCTCATCGGGGGCCACAGGCAATCTC 181	
OY 894	cggggggtgttaagtatgccaaaagtatcacatgatagatcgtgtgggaagaataccaagcg 953	
Db 182	CTATATGCGATGGGCTCTAAAGCACAGATTGAGTAGATGTGCTGGGCACAATCCTTCACAG 241	
OY 954	gtlccccaccocggtygtcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1013	
Db 242	GCTGTGTGCCCCAGAGGGCCCCCAAACACCAGCTGTCTTAAGAAGCTGTGGAGGCGCTTT 301	
OY 1014	ctccctaaccatcgtagtactagttatcaacttsgtctatgggcgggacctgtgccaaca 1073	
Db 302	CACCAATCGCTGTGTGGTTGGATTGGCCATTGCCATTCACATCGGGGAAGATTTGGCCCTGAG 361	
OY 1074	gcacgagctacgaacgtgtatltcgacaagagatgtatcgcttcgcgtcagcaacttctt 1133	
Db 362	GCAGGGCTACCGGGGTGGAGACAGCAACAGGAGCTGTGGTGGGCTCAGTAACCTTAT 421	
OY 1134	tgtgtcctctttaaatatcatatgtcaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1193	
Db 422	CGGAGGCAATCTTTCACAGTGTCTCCCGTGAAGTGTCTATGTCTCGAGAGCTGGTACAGGA 481	
OY 1194	tggagcttggaagaataatcccaagtggtgcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1253	
Db 482	GAGGACCCGGGGCAACTCGCAGGTTGCTGTGAGCAATCTTCTCTTTTATCTCTCTCAT 541	
OY 1254	catgtctgtctcgttggatctatctgtlatcctctccctaagtctgtgtgtgaagccctgtat 1313	
Db 542	CATTTCATAAACTTGGGGAACTTTCATCATGACCTGCCCAGAGGGGTGCTTGACAGCATCAT 601	
OY 1314	cgccttcaacttccaagaactccctcaaagaataacggaccctactactctgttggaggaa 1373	
Db 602	CATTGTGAACCTGGAAGGGCATGCTGTGGAGCTCAAGCGACATACGCTCTCTGTGGAGGC 661	
OY 1374	gagcaagcttgagcttgttgcattcgtgttagttagcttccctcctccttcttccatgaagct 1433	
Db 662	CAATCGGGGGATCTGCTTATCTGTGGTGTGAACCTT-CACGGGACCATCTTGCTGAACCT 720	
OY 1434	gccctatggtgtgtgcagtggtgtgtgcgtctctcgtctcgtgtgtgtgtgtgtgtgtgtgt 1490	

Db	721	GGACCTTGGCTTGGGTCGCGTCATCTCTCCCTGCTGCTCGGTCGCGTCCGAG	777
RESULT	4		
LOCUS	BF358998/c		
DEFINITION	BF358998	386 bp	mRNA
ACCESSION	CM1-ET0042	130600-264	h04 ET0042 Homo sapiens CDNA, mRNA sequence.
VERSION	BF358998.1	GI:11318070	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 386)		
	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. P., Matsukuma, A., Bala, G. S., Simpson, D. H., Brumstein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags.		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7)	3491-3496 (2000)
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL {http://www.ludwig.org.br/scripts/gethtml2.pl?rl=CM1-ET0042-130600-264-h04&f3=2000-06-13&f4=1} Seq primer: puc 18 forward High quality sequence start: 17 High quality sequence stop: 96. Location/Qualifiers		
FEATURES	source		
	1..386		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="ET0042"		
	/note="Stage="Adult"		
	/note="Organ: lung_tumor; Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	133 a	92 c	92 g
ORIGIN			69 t
Query Match	8.3%	Score 188.6	DB 11; Length 386;
Best Local Similarity	96.3%	Fred. No. 6.8e-31;	
Matches 246; Conservative	0;	Mismatches 4;	Indels 5; Gaps 5;
QY	1073	accagcgctacgaactgtatctgaacacgaagatatacgctcgtgcgcgcgaactct	1132
DB	382	acgacggctacgacgtgacgtgacgaacacgaagatatacgctcgtgcgcgcgaactct	323
QY	1133	ttggtctcctcttaaaatcgaatgcatttgcgtgtgcgttc-tgtcaacttgagctgtg	1191
DB	322	ttggctcctctttttaaattcatgtcatgttgcgtgtgcgtttctgtctgctgctgtg	263
QY	1192	gatgaagcttgaggaataatccaggttgccacc-c-tgtgtgtgtctcttggtgtgatg-a	1249
DB	262	gatgaagcttgaggaataatccaggttgccacccttgtgtgtgtctcttggtgtgatg-a	203

QY	1250	tsaactgctgagtcctggagatctctgtaaccctcc-aaagtctgtg-ctggagac	1307
Db	202		
QY	1308	cctgacgcctgcaaa	1322
Db	142		
Db	142	CCTGATCGCTGCTCA	128
RESULT	5		
AKO18566			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			

URL:<http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN,  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGACGAGATCTTCGAGTAAATTAAATTATCCCCCCCCC 3']. cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot -10.0 and subtraction to  
Rot =183.2. Second strand cDNA was prepared with the primer  
adapter of sequence [5'  
GAGACGAGATCTTCGAGTAAATTAAATTATCCCCCCCCC 3']. cDNA was cleaved  
with BamH1 and XhoI. Vector: a modified pBluescript KS(+) after  
bulk excision from lambda FliC I. Cloning sites, 5' end: SalI; 3'  
end: BamH1. Host: DH10B.

Location/Qualifiers

source  
1..1246  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1904090"  
/db\_xref="MGD:MGI:1918793"  
/clone="9030623B18"  
/sex="male"  
/tissue-type="colon"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
239..1069  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB31279.1"  
/db\_xref="GI:12858334"  
translation="MMNPDRCTPRLERSLRPRDTTRHSMTSKGVASWLPARKI  
KEMLLSDIVSGISTGLAVADLQGLFALLINVPAYGILAFPIITTFELGTSHSY  
GPFPVLSSMVGVTVRVSDPNASSLSSTSDSETEERVMVAASVTVLSIQLL  
LGVLQGVFTYVLESLSIGFTPTAAAHIVAVSOLQPMQLPVPAVSDFPSIFVLESV  
PTIOKTMIAIDLTVSIIILVVVFVEKEINORYRSKLPLPIPILIMLVIIYRSLIETI"

CDS

BASE COUNT 330 a 298 c 271 g 347 t

ORIGIN

Query Match 7.9%; Score 179.2; DB 12; Length 1246;  
Best Local Similarity 55.4%; Pred. No. le-28;  
Matches 393; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

Dbb 171 tggtctcctcctgctgcctcccaagtaacaagaattaaagactacataccttcgaagctcg 230  
340 ttgtagatcttttggtggcagacctaacacatatgaagaataggcttgcagtgcacatgcgtc 399  
231 tggactcaagcgggggatccatccaagtgctccacaagaagcatgcatctgctcgctggccaa 290  
400 tggcattcagaccggcgctggctgctgctacgctcgaagagtttacacattgctgcgtgctcaa 459  
221 cccttcctgaatcaatgctcctctaactcctctctcccctccctgaactactcctct 350  
460 catccctccagccttagcggttgtatgacgacgccctttttcccaagtcatatacctactttttctt 519  
351 gagggggtcttccacagatagtgccagcagfaccttgcgcttatacgaaccctggtyggtaa 410  
520 gggcacacatctagacacacatattcgtgggccaatttccatttcgagatgagcggggagct 579  
411 catctctcgcagctggccccagagtcgaaaatcccaagtlcttccaacatgcaccacaatga 470  
580 tctcgtttcaagagagtgctgcgaccccaa-----tgctttcttcagagattgcttca---- 631  
471 gagcatatgtgacacagcacagcattgagagctgagagagctgcaagctgtacgtacgtacgc 530  
632 -agctccacagaaatgactcgtttatagacgacgaagaaatggtggctgcatcagtataac 690  
633 -agctccacagaaatgactcgtttatagacgacgaagaaatggtggctgcatcagtataac 690







```

Db      544 ATCAATCAGCGCTCACAGACAACTTCGGTGCCCATCCCATGCATCATCATGTA 603
Oy      874 gtg 876
        ||
Db      604 TTG 606

RESULT  9
LOCUS   BG680597          611 bp      mRNA          EST          01-MAY-2001
DEFINITION 602629036F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753952 5',
            mRNA sequence.
ACCESSION BG680597
VERSION   BG680597.1 GI:13911994
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 611)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
            Email: cgapds-remail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: L1AM10614 row: 9 column: 09
            High quality sequence stop: 610.
            Location/Qualifiers
                source          1..611
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="IMAGE:4753952"
                /clone_1ib="NCI_CGAP_Skn4"
                /tissue_type="squamous cell carcinoma"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: Skin; Vector: PCMV-SPORT6; Site:1: NotI;
                Site:2: SalI; Cloned unidirectional. Primer: Oligo dT.
                Average insert size 1.5kb. Library constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      139 a      163 c      128 t
ORIGIN
Query Match          5.9%; Score 132.8; DB 11; Length 611;
Best Local Similarity 59.6%; Pred. No. 1e-18;
Matches 224; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Oy      1265 tggggatctatctgtatccttccttaagtctgtgtagagacctatgcgtcaatc 1324
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1 TGGGGAACTCTTCCATGACCTGCGCAAGCGGCTCTGGCAGCATCATGTGAACC 60
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1325 tcaagaactcccaagaactcagaccctactcctcgttgaggaagaagaagctcg 1384
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      61 TGAAGGGCATGCTGAGCACTCAGCGACATCGGCTCTCTGGAAGGCCAATCGGGCGG 120
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1385 acgtgtgcatctggtagtagtcttctcctcctcctcctcctcctcctcctcctatg 1444
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      121 ATCTGCTATATCTGGCTGTGATACCTTCACGGCCACCAATCTTGCTGAACCTTGACT 180
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1445 tggcagtggtgtgcgcctcgcctcgtcgtgctgttcacgaactcagtttcgaatg 1504
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      181 TGGTGGTGGGCTCATCTCTCCCTGCTCGTGGTGGTGGTCCGACACAGATGCCCCACT 240
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1505 gctatgcctgagcccaagtgatagcaactgacatttgaatcccaagaagccataata 1564
        || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      241 ACTCTGCTCTGGGGCAGGTGCGACACGGATATTATTACAGAGATGTGGCAGAGTACTCAG 300

```

```

Oy      1565 gggcccgatataccagggaattaaatcatcacgtactcctccctctacttgcga 1624
        ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      301 AGCCCAAGGAAGTCCGGGGGTGAAGCTCTCCGCTCGGCGCACCGTACTTGCCA 360
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1625 actcagagatcctcag 1640
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      361 ATGCTGAGTTCTTACAG 376

RESULT  10
LOCUS   BE156730          425 bp      mRNA          EST          21-JUN-2000
DEFINITION OV0-HT0368-310300-181-f03 HT0368 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE156730
VERSION   BE156730.1 GI:8619451
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 425)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
            20202663
JOURNAL   Contact: Simpson A.J.G.
            MEDLINE Laboratory of Cancer Genetics
            COMMENT Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICK Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6tz-qv0-HT0368-310
            300-181-f03&t3=2000-03-31&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 425.
            Location/Qualifiers
                source          1..425
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1ib="HT0368"
                /dev_stage="Adult"
                /note="Organ: head,neck; Vector: puc18; Site:1: SmaI;
                Site:2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."

BASE COUNT      77 a      119 c      129 g      100 t
ORIGIN
Query Match          5.7%; Score 129.2; DB 10; Length 425;
Best Local Similarity 59.6%; Pred. No. 5.7e-18;
Matches 218; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Oy      1275 tctgtatcctcctccttaagtctgtcgtagagccctgctgctcaatcctcaagaactc 1334
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      8 TCTTCATGACCTCGCCCAAGCGGCTCTGGCAGCATCATGTGAACCTGAAGGGCAT 67
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      1335 cctcaagaactcagccctactactctgttgaggaagaagaagctgagctgttgc 1394
        || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Df	68	GCTGAGGCACGTCCAGCAGACATCGCTGCCCTTGGAAGGGCAATTGGGGGAGATCTGCTTAT	127
OY	1395	ctgggttagtgagcttcctcctcctccttccctcaagctcgccctatggtgycaggtgg	1454
Df	128	ctggcttgaggacctttacagggccaccacattcttgtaacctggacgacttggctgtgc	187
OY	1455	tgtgcctctcctcgctcgtgctgctgtctccagaactgaattcgtaaagtgcgaact	1514
Df	188	GGTCATCTTCTCCCTCCCTCCTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	247
OY	1515	ggcccagatcatgatgaactgacatttatgtgaatcccaagacctataataggccagga	1574
Df	248	GGGCGAGGTGGCCAGCACAGCATATTTACAGAAATGTGGCAGTAGACTCAGAGCCACA	307
OY	1575	tatccaggagatataaatcatcacgtactgcgccctcctcctaattccaaactagat	1634
Df	308	AETCCGGGGGGGTGAAGGTCTTCCGCTCCTCGGCCACCCTGTACTTGGCCAATGCTGATT	367
OY	1635	cctcag 1640	
Df	368	CTACAG 373	
RESULT	11		
BB605137			
LOCUS	BB605137	283 bp mRNA	EST 05-DEC-2000
DEFINITION	BB605137 RIKEN full-length enriched, 0 day neonate lung Mus musculus cDNA clone E030008K16 5', mRNA sequence.		
ACCESSION	BB605137		
VERSION	BB605137.1	GI:11556539	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Alizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Koijima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sekuri,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watabiki,Y., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshihi,A., Muramatsu,M. and Hayashizaki,Y.		
TITLE	RIKEN Mouse ESTs (Alizawa,K. et al. 2000)		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shuhito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermosensitization and thermocatalysis of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,K., Ozawa,T., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.rtc.riken.go.jp">http://genome.rtc.riken.go.jp</a> ) for further details.		
FEATURES			
SOURCE	1..283	Location/Qualifiers	

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E030008K16"
/clone_id="RIKEN full-length enriched, 0 day neonate
lung"
/tissue_type="lung"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/notes="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAGCGCGCCGACACGAGCTTTTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGGAGAGAGATTCTCGAGTTAAATTAATTAATCCCCCCCCCCC 3']. cDNA
was cloned with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLX I."

```

Query Match	5.6%	Score 125.8	DB 10	Length 283
Best Local Similarity	80.0%	Pred. No. 2.8e-17		
Matches 148	Conservative	0	Mismatches 37	Indels 0
				Gaps 0

Oy	1	atagagccagccagcagcccgctcgaagcttgtaagaagaacgcgaatactccctaccccttc	60
Db	87	ATGAGCCAGGCCAGGCCCGCTGACGTGGTAATATAGAGTGGATATCCCTCTCCTCTTC	14
Oy	61	gacgatagcttgaagaagaagaccggagacataccagctgaggaagaacttcgcaatgcc	12
Db	147	GATGATATATTTTGAAGAAAAGAGATCGAGCCCTACCCGGGGAGAGAAAGCTTCTTAACT	20
Oy	121	tccagatgcttcctcagcccaagaatcaaacgctgctggttggctctgctgctctgctctcc	18
Db	207	TCCGGGTGCTCTCTCACCCCAAGTTCAAAACCTATGTGTGGGGGTGTGGCTGTGCTCTCC	26
Oy	181	tggct 185	
Db	267	TGGCT 271	

RESULT	12				
LOCUS	BG732794				
DEFINITION	BG732794	521 bp	mRNA	EST	11-MAY-2001
ACCESSION	364620	MARC	1Pig Sus scrofa	CDNA 5',	mRNA sequence.
VERSION	BG732794				
KEYWORDS	EST.				
SOURCE	EST.				
ORGANISM	pig.				
REFERENCE	Sus scrofa				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 1 (bases 1 to 521)				
	Fahrenkrug, S.C., Treking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E. Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.				
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine				
JOURNAL	Unpublished. (2000)				
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smtlth@mail.marc.usda.gov				
	Single pass sequencing. Bases called and alt-trimmed with phred				



QY 1585 altaaatcaatcagctactcctccctctactcttgccaaatcagagatcttcag 1640  
 Db 359 GTGAAGCTTCCTCCTCCTCAGCCACGCTGACTTCGCAATGCTGAGCTCTACAG 414  
 RESULT 14  
 LOCUS BG727939 558 bp mRNA EST 09-MAY-2001  
 DEFINITION fo79h02.y1 zebrafish gridded kidney Danio rerio cDNA clone 4726491  
 5' similar to SW:DR1\_HUMAN P40879 DRA PROTEIN ; mRNA sequence.  
 ACCESSION BG727939  
 VERSION BG727939.1 GI:14013014  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Rasbortinae; Danio.  
 REFERENCE 1 (bases 1 to 558)  
 AUTHORS Clark,M., Johnson,S.L., Lebrach,H., Lee,R., Li,F., Marra,M., Eddy  
 S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
 K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
 Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
 and Wilson,R.  
 TITLE Washu zebrafish EST Project 1998  
 JOURNAL Unpublished (1998)  
 COMMENT Other ESTs: fo79h02.x1  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zdraflsh@watson.wustl.edu  
 CDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:  
 Washington University Genome Sequencing Center Clone Distribution:  
 Genome Systems, St. Louis, Missouri (web address:  
 www.genomesystems.com) (email contact: info@genomesystems.com) and  
 Research Genetics, Huntsville, Alabama (web address: www.resgen.com  
 ) (email contact: info@resgen.com) and  
 ResourcenzentrumPrimarDatenbank, Berlin, Germany (web address:  
 www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 488.  
 FEATURES  
 source location/Qualifiers  
 1..558  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone\_id="4726491"  
 /clone\_lib="zebrafish gridded kidney"  
 /sex="mixed"  
 /tissue\_type="kidney pooled from 300 wild type adults"  
 /lab\_host="XLOLR"  
 /note="Organ: kidney; Vector: pBR-CMV; Site:1: EcoRI;  
 Site\_2: XhoI; Oligo dT cDNA library constructed from mRNA  
 pooled from pooled kidney tissue from 300 adult  
 zebrafish."  
 BASE COUNT 126 a 120 c 141 g 171 t  
 ORIGIN  
 Query Match 5.4%; Score 121.8; DB 11; Length 558;  
 Best Local Similarity 51.6%; Pred. NO. 2.5e-16;  
 Matches 279; Conservative 0; Mismatches 262; Indels 0; Gaps 0;  
 QY 1086 aaccagagatgatcgtctcgtcgtcgcagcaattcttgctcctcttaaaattat 1155  
 Db 4 AATAGAGGCTCATCTCGCTGGACTTGTACTTGTGCACATCTCTTCACACGTTT 63  
 QY 1156 gtcaattgctgtcgcttcctgcacactcgtgctgtgagtgagctgagagaaatccag 1215  
 Db 64 GTCTGACCGCTCAATCTCTCGCAGCTGTGACAGAGACATTGGAGACATACTGAA 123

QY 1216 gtggcagcctgtgtgtgtctctgtgtgltgaltcaccatcgtgtccctggagatcat 1275  
 Db 124 ATTGACGATCTCTTGGCTCCCTACTTGTCTGCTACTGCTGAGTGCGCATTTGATTGTC 183  
 QY 1276 ctgtatcctctccctaaatcgtgtgtgtagagagcccgatcgtctgtaactcgaagacc 1335  
 Db 184 TTTCAGCCATTACCCACACAGCGTGTGGCTGTATCATCATGTTAAATCTCTGGGATG 243  
 QY 1336 ctcaagcaactcagaccctcactacactgtgtggaagaagacagctgactgtgcatc 1395  
 Db 244 TTTAAGCAACAGAGGATATTCCAGTTTATGAGCAAAAAACAGAGATTGAACTGGCATT 303  
 QY 1396 tgggtagtgtgactctctcctcctcctcctcctcagcgtccctatgltgtgtgagtggt 1455  
 Db 304 TGGCTGTGTCTTTTGTTCATTCGCTTCTCCTGGCCCTGAGTTTATGCTTGGCGTTC 363  
 QY 1456 gtgcctctcctcgtctcgt 1515  
 Db 364 ATGCTTTTGTGCTATACCTACCGTGTATTCAGAACACACGCCCCAGCAATGCTGTTCT 423  
 QY 1516 gccacgactgagacactgacattatgtgaatcccaagacctataataggccagat 1575  
 Db 424 GGACAAATCCAGATACAGAGATTGTGACTTGTGATGTGAATATGAGAGCGTGGAGAA 483  
 QY 1576 altcagggatlaaataatcagctactgtccctcctcctcctcctcctcctcctcctc 1635  
 Db 484 TGCTTAGGATTAAGATTTCACGTCATTCCTCATCTACTCCGCAACAGAGAGCTG 543  
 QY 1636 t 1636  
 Db 544 T 544

RESULT 15  
 LOCUS AA871419 510 bp mRNA EST 16-MAR-1998  
 DEFINITION vq35b04.f1 Barstead bowel MRLB9 Mus musculus cDNA clone  
 IMAGE:1096207 5' similar to SW:DR1\_HUMAN P40879 DRA PROTEIN ; mRNA  
 sequence.  
 ACCESSION AA871419  
 VERSION AA871419.1 GI:2966864  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 510)  
 AUTHORS Matsui,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Getzel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The Washu-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:602439  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 468.  
 FEATURES  
 source location/Qualifiers  
 1..510  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1096207"  
 /clone\_id="Barstead bowel MRLB9"







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 12:01:00 ; Search time 98.36 Seconds  
(without alignments)  
5208.349 Million cell updates/sec

Title: US-09-749-589-1

Perfect score: 2262  
Sequence: 1 atgagccagccagccagccgcg.....agaccctgacgcctgtga 2262

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCrus.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	326.2	14.4	2882	1 US-08-424-567-1	Sequence 1, Appli
2	326.2	14.4	2882	2 US-08-711-928-1	Sequence 1, Appli
3	326.2	14.4	2882	4 US-09-184-937-1	Sequence 1, Appli
4	120	5.3	588	4 US-09-385-982-129	Sequence 129, App
5	111.4	4.9	623	4 US-09-385-982-513	Sequence 513, App
6	106.2	4.7	475	4 US-09-385-982-146	Sequence 146, App
7	105.2	4.7	595	4 US-09-385-982-477	Sequence 477, App
8	67.4	3.0	583	4 US-09-385-982-491	Sequence 491, App
9	53.8	2.4	7218	1 US-08-237-463-14	Sequence 14, Appli
10	48.4	2.1	816	2 US-07-637-865-1	Sequence 1, Appli
11	48.2	2.1	3489	2 US-08-728-323A-1	Sequence 1, Appli
12	48.2	2.1	32207	2 US-08-770-379-20	Sequence 20, Appli
13	48.2	2.1	32207	4 US-08-757-669A-20	Sequence 20, Appli
14	46.8	2.1	1931	2 US-09-130-114-2	Sequence 2, Appli
15	45.2	2.0	735	4 US-08-998-416-211	Sequence 211, App
16	44.2	2.0	3061	1 US-08-700-576-1	Sequence 1, Appli
17	42.4	1.9	1465	4 US-09-338-671-1	Sequence 1, Appli
18	42	1.9	4403765	4 US-09-103-840A-2	Sequence 2, Appli
19	41.4	1.8	2220	2 US-08-617-801A-1	Sequence 1, Appli
20	40.2	1.8	289	4 US-09-007-005-17	Sequence 17, Appli
21	40.2	1.8	289	4 US-09-244-796-17	Sequence 17, Appli
22	40	1.8	4403765	4 US-09-103-840A-2	Sequence 2, Appli
23	39.8	1.8	1189	1 US-07-781-034-4	Sequence 4, Appli
24	39.8	1.8	1189	5 PCT-US92-08328-4	Sequence 4, Appli
25	39.2	1.7	2712	3 US-09-025-691-4	Sequence 1, Appli
26	39	1.7	1776	1 US-08-484-840-1	Sequence 1, Appli
27	39	1.7	1776	1 US-08-483-094-1	Sequence 1, Appli

28	39	1.7	1776	4 US-09-318-448-13	Sequence 13, Appli
29	38.8	1.7	3300	2 US-08-194-290-6	Sequence 6, Appli
30	38.8	1.7	3300	2 US-08-614-377A-6	Sequence 6, Appli
31	38.8	1.7	3300	4 US-09-142-648B-6	Sequence 6, Appli
32	38.6	1.7	1722	4 US-09-385-028-15	Sequence 15, Appli
33	38.6	1.7	11604	4 US-09-385-028-13	Sequence 13, Appli
34	38.6	1.7	15079	4 US-09-385-028-1	Sequence 1, Appli
35	38	1.7	36519	3 US-08-923-137-2	Sequence 2, Appli
36	37.8	1.7	1491	4 US-09-082-092-9	Sequence 9, Appli
37	37.6	1.7	7295	2 US-08-487-826B-15	Sequence 15, Appli
38	37.4	1.7	20235	1 US-07-642-734C-3	Sequence 3, Appli
39	37.4	1.7	20235	3 US-08-439-009A-3	Sequence 3, Appli
40	37.2	1.6	966	2 US-08-766-738-2	Sequence 2, Appli
41	37.2	1.6	1227	3 US-09-074-912-3	Sequence 3, Appli
42	37.2	1.6	9960	3 US-08-822-586-46	Sequence 46, Appli
43	37	1.6	2156	2 US-08-899-514-1	Sequence 1, Appli
44	37	1.6	3318	4 US-09-593-589-3	Sequence 3, Appli
45	37	1.6	15872	4 US-09-105-537-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-424-567-1  
; Sequence 1, Application US/08424567  
; Patent No. 5569755  
; GENERAL INFORMATION:  
; APPLICANT: SCHWEINFEST, Clifford W.  
; TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,567  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/026,045  
; FILING DATE: 05-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/181 NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2882 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 185..2479  
; US-08-424-567-1  
Query Match 14.4%; Score 326.2; DB 1; Length 2882;



```

: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2882 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 185..2479
: US-08-111-928-1

```

[illegible]

Db	1094	TCCTACGCGCTGTGACTTTAAACAAAGSTTTAAAGTGGCTGTGGTGGGACATGATTCCT	115
Qy	952	gggttcccccaccccggtgcgcctgtgtgttcacagltgaaagacatgatagaccagcc	101
Db	1154	GGATTTTCAGGCCCTATTTAACACCTGACGTGGAGACTTTCCAAAACACCGTAGAGATTGC	121
Qy	1012	tttccctcctagacatcgcgtagctacgtcatcaactgtgtctatgggccggaacctggccaac	107
Db	1214	TTTCGGCATCGCAATGGATTTGGCATTTGGCAGTGGCCCTTTTCAGTTCCACCGCTCATTTCCCTTC	127
Qy	1072	aagacagcgtctacgcagctggtattcgaaccagagatgtagctgcctgcgtcgaacacttc	113
Db	1274	AAATACGATTATTCACCTTTGATGGCAATCAGAGAGTTATTAACCTTTGGAGCTGGGTAACTTA	133
Qy	1132	tttgcctcctctctttaaataatcatgctatttgcctgtgtgcgtcttcctgtcaactgtgcgtg	119
Db	1334	GTCGTGGAGATTAATCAGAGAGATTGCTGGGAGTACTCCCTCTCCACATGACGATTTTCAG	139
Qy	1192	gatgtgagctgtgaggaataatcccaagtggtccagcctgtgtgtgtctctcgtgtgtgagatc	125
Db	1394	GAGAGCACAGAGAGCAAAACACAAATTTGCTGGGCTCATTTTGGTGGCCATCATCGCTGTGATT	145
Qy	1252	acacatcgtgtgcctcggggagatctatcgtctatccctccctcctaagctcgtgtcgaagccctg	131
Db	1454	GTCGTTCTAGGCATTGGATTTTCTCTGTGGCCCTCTACAAAAGTCCGCTCGGACGCTTTA	151
Qy	1312	atcgcgttcaactctcaagaaactccctcaagaacatccagaccctactaactctgttgag	137
Db	1514	GCATTTGGGAACCTTAAGAGGAATCGTATGACGATTTGCTGCAAAATAGCAGATTTGTGGCA	157
Qy	1372	aagagcaagctgtgacgtgtgtcatctgtgtgtagtgtgactctctcctcctcctccagc	143
Db	1574	AAGGACAATATATGATTTGATTAATTTGGATCATGACCTTCATCTTCCACCATTTGCTGGGA	163
Qy	1432	ctgcgcctatgtgtgtgagcagtggtgtgcgccttcgcctccgtccctgtgtcgtgtctccagact	149
Db	1634	CTCGGCTTAGGCCCTGCGCAGCTAGTGTGGCATTTTAACCTGCTAACCATCGTGTTCAGSACC	169
Qy	1492	caatttcgaaatgtgctatgcatgtgcctggccaggtcatgtgacactgtgacatttgtgaatccc	155
Db	1694	CAATTTCCAAAATGCAGACAGCCTGGCTAATATTGGAAGAACCAACATCTAATGAATATAA	175
Qy	1552	aagacataatagggcccaagatatccagagggaatataatcatcaactgcgtcctccct	161
Db	1754	AAAATATATATATGATATGATGTAAGACCACGAAGAGTGAATAATTTTCACATGTCCATCTCCT	181
Qy	1612	ctctactcttgcgaactcagagatcttcaagaaaggtaatcgc	1654
Db	1814	ATCTACTTTGCAAAACATTGGTTCTTTTAGCGGGAACATTATCG	1856

RESULT 3

US-09-184-937-1

Sequence 1, Application US/09184937

Patent No. 6210887

GENERAL INFORMATION:

APPLICANT: SCHWEINERT, Clifford W.

APPLICANT: PAPAS, Takis S.

TITLE OF INVENTION: Colon Mucosa Gene Having Down-regulated

TITLE OF INVENTION: Expression in Colon Adenomas And Adenocarcinomas

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25



```

1 GENERAL INFORMATION:
2 APPLICANT: ENDEGE, WILSON O., ET AL.
3 TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
4 TITLE OF INVENTION: PRODUCTS: II
5 FILE REFERENCE: CCDNA-260XX
6 CURRENT APPLICATION NUMBER: US/09/385,982
7 CURRENT FILING DATE: 1999-08-30
8 EARLIER APPLICATION NUMBER: 09/328,111
9 EARLIER FILING DATE: 1999-06-08
10 EARLIER APPLICATION NUMBER: 60/117,193
11 EARLIER FILING DATE: 1999-01-27
12 EARLIER APPLICATION NUMBER: 60/098,639
13 EARLIER FILING DATE: 1998-08-31
14 NUMBER OF SEQ ID NOS: 544
15 SOFTWARE: FastSeq for Windows Version 3.0
16 SEQ ID NO 129
17 LENGTH: 588
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: (1)..(588)
23 OTHER INFORMATION: n = A,T,C or G
24 US-09-385-982-129

```

```

? CURRENT FILING DATE: 1999-08-30
? EARLIER APPLICATION NUMBER: 09/328,111
? EARLIER FILING DATE: 1998-06-08
? EARLIER APPLICATION NUMBER: 60/117,393
? EARLIER FILING DATE: 1999-01-27
? EARLIER APPLICATION NUMBER: 60/098,639
? EARLIER FILING DATE: 1998-08-31
? NUMBER OF SEQ ID NOS: 544
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 513
? LENGTH: 623
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: (1)..(623)
? OTHER INFORMATION: n = A,T,C or G
US-09-385-982-513

```

Query Match	5.3%	Score 120:	DB 4:	Length 588:
Best Local Similarity	53.1%	Fred. No. 9.7e-21:		
Matches 255:	Conservative 0:	Mismatches 221:	Indels 0:	Gaps 0:
Oy	1167	tgcgcttcctgtcagcatcvgctgctgaltgagctgtagaggaatacccaagtgcgcgacct	1226	
Db	6	tgcccttcocagatcaagcaagttcaagagagcaccaggaagcaaacacacagatgtcctggagct	65	
Oy	1227	gtgtgtgtctctgtgtgtgtatgataccaatcgtgtcctgtggagatctatgtatccct	1286	
Db	66	tattgtgcacatcactgtgtgtatgttcgtcttaagcattgttatctctcctgtgcgacct	125	
Oy	1287	ccctaagctgtgtgtcagagagccctgtatcgtctcaatctccagaagctccctcaagaact	1346	
Db	126	acaaaagtcgcgtccctggcagcgtcttgatgtgaaactaaagggagatctbatgcagtt	185	
Oy	1347	cacgcagccctactactcctgtgtgaggaagaagcagctgtgacgtgtgcattgtggtatgag	1406	
Db	186	tgctgaatataggcagattgtggcgaaagacaaatatgattgtttaattgatcatgac	245	
Oy	1407	ctctctctccctctctctctcagcctgcgcctatgtgtgagagtggtgtgcctctcc	1466	
Db	246	ctctactctcaccaattgtcttcgtgagctcgtgttaaggccggacagctagtgtgcattca	305	
Oy	1467	cgcttcgtgtgtgtgtccttcacagatcagttctgaaatgtgtatctgtcgtggccaggtcat	1526	
Db	306	actgttaaccactcgtgtctcagagaccaatcttcaaaatgcagcagcgtgcataattg	365	
Oy	1527	ggacctgcacattatgtgaaatcccaagaccataataggcccgagatataccaggagat	1586	
Db	366	aagaaccacacataatgaataaaaaaagattatattgatatgtatgagccagaaagagt	425	
Oy	1587	taaaatcatcactgtactgtccctctctactcttgcacactagagatctcaggaaaa	1646	
Db	426	ggaattttcagatgtccctcctcactctcactcttgcaaaacattgtttctttaggsggaa	485	

RESULT 5  
 US-09-385-982-513  
 ; Sequence 513, Application US/09385982  
 ; Patent No. 6262334  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ENDEGE, WILSON O., ET AL.  
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
 ; FILE REFERENCE: CCDNA-260XX  
 ; CURRENT APPLICATION NUMBER: US/09/385,982

```

Query Match Similarity      4.9% Score 111.4 : DB 4 : Length 623 ;
Best Local Similarity      52.9% : Pred. No. 1.3e-18 ;
Matches 258 ; Conservative 0 ; Mismatches 225 ; Indels 1 ; Gaps 1 ;

QY 1167 tgcgcttctcgtcaatccgtcgctgtgatagtcgtgaggagaataaccagtgtagcacgct 1226
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 tgcctctccagatcagcagttcagagagacacagaggaacaacaacagattgcgggct 62
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1227 gtgtgtctctgtgtgtgtatgaatcaacatgctgtgctcgtggatcatctgttatcct 1286
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 63 tatgtgtgccatcatcgttgtatgtctgtcttaagcattgtattctctcgtgcacct 122
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1287 ccctaagtctgtgcttagagagccctgtatctgcatactccaagaactccctcaagaact 1346
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 123 acaaaagctgcctccgycgaccttagcatctggaaaacttaaaaggaaatgcgtatgcgtt 182
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1347 caaccacccctactaacctgtgaggaaagaaagacaagctgtgttgcatactgtgtatgag 1406
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 183 tgcigtaaaaagcagattgtgtgagaaagacaacaatatgttctaatgttgcataagac 242
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1407 ctctctctctctctctctctctcctcagcctgcctatgtgtgtgcagtgtgtctctctc 1466
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 243 cttcatcttcaacattgtctctgtgagctcggtgttaaggccgtgaagtgttgcatttca 302
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1467 cgctcgtgtcgtgtgtcttccacacacagttcgaataatgtgactatgcagccaggtcat 1526
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 303 actgctnaacacatcgtgttcaaggacccaattcttcaaataatgcagacgtggtcaaatgtg 362
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1527 ggcacatgcatacttatgtgaatcccacagcctataaaggccagatcatcaggagat 1586
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 aagaacccaacatcataaagaataaaaaagattatcatgtatgtatgtatgacgaagagat 422
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1587 taanaatcatcgtactgtctccctctacttctacttgcacaaclcagasatcttaagcaaaa 1646
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 423 gaanaat-ttcagatgtccatctctcatctactactgtgnaacattgnttctttagcngga 481
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1647 ggctcatcg 1654
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 482 acctatcg 489
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT
US-09-385-982-146
; Sequence 146; Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDA--260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111

```

```
;; EARLIER FILING DATE: 1999-06-08
;; EARLIER APPLICATION NUMBER: 60/117,393
;; EARLIER FILING DATE: 1999-01-27
;; EARLIER APPLICATION NUMBER: 60/098,639
;; EARLIER FILING DATE: 1998-08-31
;; NUMBER OF SEQ ID NOS: 544
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 146
;; LENGTH: 475
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(475)
;; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-146
```

```
Query Match      4.7%, Score 106.2; DB 4; Length 475;
Best Local Similarity 56.5%; Pred. No. 2.4e-17;
Matches 262; Conservative 0; Mismatches 189; Indels 13; Gaps 3;
```

```
Qy 259 ccacagagcatgagcattgctctgtgtgcacaccccttcctgcagtaatggtccctacc 318
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2 ctacaaggtttagcaattgtctgtctgtgcacattcccccagttctatggtgtatgca 61
Qy 319 tcccttcccccctcctgaactactctcctctggtgggtgttccaccagatggtgccagt 378
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62 tccctttcccaagcataatctactcttctctgcgaccttccagacacatatacgttggt 121
Qy 379 acccttgcggtatcagaatcctgtgtgg-----taacatctgtctgcagctggt 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 ccgttccgattctgatatgtagtggactagcagtttcaggagcagtttcaaaagca 181
Qy 428 cccacagatggaattccagagctctcaacaatgccc-accacagagagcatatgtgacaca 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 gtccacagatgcgaatgcgaactactcttggatgtgcttaacaactgaaatcttcaacta 241
Qy 487 gcagcagatgagagctgtagagagctgcacgtgtcagctacgctgcctcacccgcatc 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 ctgagtgacgagaggttgaggttggcggcgcgacatcagtcacagttgtcttcgtgaatc 301
Qy 547 atccagatgaggtctggtcttcatgtcagtttggcttggcgcatctactctccgagtc 606
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 302 atccagttggtcttggatctcgtgattgtgattgatatataacctgtctgagttc 361
Qy 607 ttcaatccgggcttcatgacgcccgcgcgcgcagatccattcgtgtgctcaagtaac 666
    ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 362 ctcaatcagttgcttcaactacagctg-cgtcgtncatgttttggttcccaactcaatlc 420
Qy 667 atcttcgagctgacatccctcctctacacagcccaagggtccat 710
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 attttcaattgacagctccgcgtcacacacatgctcagtttcaat 464
```

```
RESULT 7
US-09-385-982-477
; Sequence 477, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
```

```
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 477
;; LENGTH: 595
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(595)
;; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-477
```

```
Query Match      4.7%, Score 105.2; DB 4; Length 595;
Best Local Similarity 56.1%; Pred. No. 4.5e-17;
Matches 261; Conservative 0; Mismatches 191; Indels 13; Gaps 3;
```

```
Qy 259 ccacagagcatgagcattgctctgtgtgcacaccccttcctgcagtaatggtccctacc 318
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2 ctacaaggtttagcaattgtctgtctgtgcacattcccccagttctatggtgtatgca 61
Qy 319 tcccttcccccctcctgaactactctcctctggtgggtgttccaccagatggtgccagt 378
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62 tccctttcccaagcataatctactcttctctgcgaccttccagacacatatacgttggt 121
Qy 379 acccttgcggtatcagaatcctgtgtgg-----taacatctgtctgcagctggt 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 ccgttccgattctgatatgtagtggactagcagtttcaggagcagtttcaaaagca 181
Qy 428 cccacagatggaattccagagctctcaacaatgccc-accacagagagcatatgtgacaca 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 gtccacagatgcgaatgcgaactactcttggatgtgcttaacaactgaaatcttcaacta 241
Qy 487 gcagcagatgagagctgtagagagctgcacgtgtcagctacgctgcctcacccgcatc 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 ctgagtgacgagaggttgaggttggcggcgcgacatcagtcacagttgtcttcgtgaatc 301
Qy 547 atccagatgaggtctggtcttcatgtcagtttggcttggcgcatctactctccgagtc 606
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 302 atccagttggtcttggatctcgtgattgtgattgatatataacctgtctgagttc 361
Qy 607 ttcaatccgggcttcatgacgcccgcgcgcgcagatccattcgtgtgctcaagta 665
    ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 362 ctcaatcagttgcttcaactacagctgtcgtctcattgttttggtttcccaactcaatlc 421
Qy 666 catcttcgagctgacatccctcctctacacagcccaagggtccat 710
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 cattttcaagtgacagctccgcgtcacacacatgatacagtttcaat 466
```

```
RESULT 8
US-09-385-982-491/c
; Sequence 491, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 491
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

LENGTH: 7218 base pairs

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 01-338267  
;; FILING DATE: 28-DEC-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Skeppon, Joseph M.  
REGISTRATION NUMBER: 29, 864  
REFERENCE/DOCKET NUMBER: 0118.033382  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-296-5500  
TELEFAX: 202-296-7830  
INFORMATION FOR SEQ ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 816 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Flavobacterium  
STRAIN: Flavobacterium sp. No. 5942427 141-8  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..813  
OTHER INFORMATION:  
US-07-637-865-1

Query Match 2.1%; Score 48.4; DB 2; Length 816;  
Best Local Similarity 53.1%; Pred. No. 0.0061;  
Matches 103; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

OY 1739 acggaccaggtgtcctatataccttcagccctgcagagctccacactgcagagtg 1798  
DB 425 AGGGCGGCTCCGGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 484  
OY 1739 agccacacacccctcgcctcagagcccccgcagagccagtgacatgctgscagcgtccac 1858  
DB 485 AGGGCGGCGCTCCGGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 544  
OY 1859 ccttcgtacactccacacacccctcagacatgagtgagtgacagcttcgtgagacttga 1918  
DB 545 TCGACCTCGCGCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 604  
OY 1919 tgggacataagggc 1932  
DB 605 CCGGCAACACACAC 618

## RESULT 11

US-08-728-323A-1/c  
Sequence 1, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/728, 323A  
APPLICATION NUMBER: US/08/728, 323A

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28, 678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Query Match 2.1%; Score 48.2; DB 2; Length 3489;  
Best Local Similarity 46.3%; Pred. No. 0.012;  
Matches 158; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

OY 534 cctcacgcgcacatcacagatggtcgtggtcctacagattggtcgtggtccatcta 593  
DB 1288 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 1229  
OY 594 ccttcagagctcctcctcagaggtcctacatgaagcgcgcgcgcgcgcgcgcgcgcgcgc 653  
DB 1228 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 1169  
OY 654 ggtgtcagaatcctcctcagaggtcctcctcctcctcctcctcctcctcctcctcctcct 713  
DB 1168 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 1109  
OY 714 ccttcacatcattgacattgcaaaacctcccccacacacacacacacacacacacacacac 773  
DB 1108 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 1049  
OY 774 cgtctcacaagcgtgctcctcgtggtggtggtggtggtggtggtggtggtggtggtggt 833  
DB 1048 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 989  
OY 834 caagattcgtcctcccccac 874  
DB 988 CCGGCTCCTCCGCGCATCATCACCATCGGCTCCGTCATTCCTTCATGCGCGAGCGG 948

## RESULT 12

US-08-770-379-20  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk









04-NOV-1998; 980S-0172255.  
PR 24-NOV-1998; 980S-0172252.  
PR 22-DEC-1998; 980S-0172214.  
PR 26-FEB-1999; 990S-0121896.  
XX  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
XX  
PI Hillman JL, Yue H, Tang YT, Lal P, Corley NC, Guegler KJ;  
PI Baughn MR, Azimzal Y, Lu DAM;  
XX  
XX WPI: 2000-365576/31.  
DR P-PSDB: AAY71067.  
XX  
XX  
PT Novel human membrane transport proteins useful for diagnosis,  
PT prevention and treatment of membrane transport disorders,  
PT immune/inflammatory disorders and cell proliferative disorders  
PT including cancer .  
XX  
XX Claim 9; Page 119-120; 136pp; English.

CC The present sequence is a cDNA encoding membran transport protein.  
CC MRP-12 from Incyte clone 17334/7 isolated from human BRST17U08 cDNA  
CC library. MRP-12 shows homology to mouse transporter (GI 5353730)  
CC and human transporter (GI 291964), and is expressed in nervous,  
CC gastrointestinal and reproductive tissues.  
CC The present sequence is useful in diagnosis, prevention and treatment  
CC of disorders related with increased or decreased expression of MRP  
CC such as familial goitre, Menkes disease, diabetes, Parkinson's disease,  
CC neurological disorders such as Alzheimer's disease, depression, epilepsy,  
CC schizophrenia, immune/inflammatory disorders such as AIDS, Addison's  
CC disease, allergies, atherosclerosis, Graves disease, gout, Hashimoto's  
CC thyroiditis, viral, bacterial, fungal, parasitic, protozoal or  
CC helminthic infections and cell proliferative disorders such as cancer.  
CC Fragments of MRP polynucleotides are useful as targets in microarrays.  
CC MRP DNA is also useful for generating hybridisation probes useful in  
CC mapping genomic sequences and detecting differences in sequences among  
CC normal, carrier and affected individuals. It is also used for  
CC screening libraries of compounds in drug screening techniques.

Query Match	17.0%	Score 384;	DB 21;	Length 2580;
Best Local Similarity	54.2%;	Prod. No. 9,3e-86;		
Matches 820;	Conservative 1;	Mismatches .670;	Indels 24;	Gaps 1;
QY 127	tgctccagcgaagaatcaaaagctgtggtgtttgtgagctgtgcctcgtgtgtctccagctc	186		
Db 215	tgctccagctgtctgggctctagcctctctgtctccaacaacctccggttlttgcgtgta	274		
QY 187	cccaagatacaagatlaaaggactacatacatctctgaacctgtctcgtgtgagactcagcgagga	246		
Db 275	ccccggtatctctgtcgtgagactgtcctctgtgagacctgtatccgcgctgagtgtygcgc	334		
QY 247	tcgaatccagggcccccaaggagatggcatttggctctcgtcggccaacctctctcagtaaat	306		
Db 335	atcatgacgcttccgcgaaggctctggcctacgccccctccggcgagatgtgccccgvtgtt	394		
QY 307	ggacctatactctcctcttctcccccctctctgaacctacttctctctcgtgggggtttaccaa	366		
Db 395	ggacctatagtcctctcttaccctgtcttcatctacttccctgttcttgcaattcccgcaac	454		
QY 367	atggtgccaagttactcttgcgcgttatcaagcaatcccggtgtggtgtaacaatctgtctcagctg	426		
Db 455	atcttcggtggggaaccttgcgtcgtcatgtctgtgagtggtggggtgtgtacagaatccctg	514		
QY 427	ggcccaaggtcgaaatctcaagctcttcaacaatgacccaatgagagctatgttgcacaca	486		
Db 515	ggcccggaaggctctggaagaactccatgtatcaatgaga-----	551		
QY 487	gcagccatlgagagctgagagagctgcaagctgtcaagctacgcgttagcctctccacgcacatc	546		
Db 552	-caagcagaagttgtgcgccgggtacaggttgagcttccacacatcaagttctcgtttgcctc	610		

[illegible][illegible]

Oy	1627	tcagagatcttcag	1640
Db	1691	gcgtgattctaacg	1704
<hr/>			
RESULT 2			
ID	AAI62799	standard; cDNA: 807 BP.	
XX	AAI62799;		
XX	22-OCT-2001	(first entry)	
DE	Human CDNA SEQ ID NO 58.		
XX			
KW	Human; nontropic; neuroprotective; cyostatic; dermatological; virucide;		
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;		
KW	antiparkinsonian; antischling; antlanemic; antlaribric; cancer;		
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;		
KW	antiallergic; antidiabetic; antituber; anticonvulsant; antifungal;		
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;		
ss.			
OS	Homo sapiens.		
XX	WO20015449-A1.		
XX	02-AUG-2001.		
PD			
PF	17-JAN-2001; 2001WO-US01346.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUL-2000; 2000US-0216880.		
PR	14-JUL-2000; 2000US-0218290.		
PR	14-AUG-2000; 2000US-0225447.		
PR	01-SEP-2000; 2000US-0229343.		
PR	06-SEP-2000; 2000US-0230437.		
PR	25-SEP-2000; 2000US-0234997.		
PR	29-SEP-2000; 2000US-0236367.		
PR	13-OCT-2000; 2000US-0239937.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246528.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249265.		
PR	01-DEC-2000; 2000US-0250160.		
PR	01-DEC-2000; 2000US-0250391.		
PR	05-DEC-2000; 2000US-0251030.		
PR	05-DEC-2000; 2000US-0251988.		
PR	05-DEC-2000; 2000US-0256719.		
PR	06-DEC-2000; 2000US-0251479.		
PR	08-DEC-2000; 2000US-0251988.		
PR	08-DEC-2000; 2000US-0251990.		
PR	11-DEC-2000; 2000US-0254097.		
PA			
XX	(HUMA - ) HUMAN GENOME SCI INC.		
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI; 2001-476225/51.		
DR	P-PSDB; AAM42394.		
XX			
PT	Novel plasma membrane associated proteins useful for diagnosing,		
PT	treating, preventing and/or prognosing disorders related to the		
PT	proteins, including cancer, immune response and neuronal disorders		
-			



```

Db 674 aataatcttcaactactgagatgacgagaggtgaggtgctgcgcgcgcgcacatcgatcaca 733
Oy 532 tgcctcaacgcgcacatccacagatggtctgggtctcatgcatggtttgttgccatc 591
Db 734 gtgtcttctggaatcaatccacatggtcttctggatctctggagatctggttgatgata 793
Oy 592 tacctctcgagatcttcaatccgggtctcatgacgcgcgcgcctgacagatctgatt 651
Db 794 taactgtctgagctccatccatcagctgctcaactgtctgtctgtctcatgttttggct 853
Oy 652 tcggtgtcaagtacatcctctggaactgacatccctctcaacagcccgaggttcac 711
Db 854 tcccaactcaaatcatcttctcagttgacagttccgttcaacacatgacatccagttcaatt 913
Oy 712 gtcttcaactcatgagatgacatgacaaacctccccacacatcgctctgtatc 771
Db 914 ttcaaatgatactatctctgtatcttcacaaatagagaagataatctgacagacctgtg 973
Oy 772 ttgcctctcaacgagtgctctctctgtgtgtggtgaagagctcaatgctctgatactg 831
Db 974 acagctctgattgtctcttggttgtgatacattgtttaagaataaatcagcgcttcaaa 1033
Oy 832 cacaagaatgctctcccatccatccacagagatgattgtgtgtgtgtgtgcaacagctatc 891
Db 1034 gacaaacttccagtgccatcttccatccatccatccatccatccatccatccatccatcc 1093
Oy 892 tccgggggagctgtaagatgacccaaagatcacatgacatgagctggtggaatcccaacgc 951
Db 1094 tcccaagcgtgtgacttaaaaaacaggtttaaagtgtgtgtgtgtgtgtgtgtgtgtgt 1153
Oy 952 gggttccccaccgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1011
Db 1154 ggaattcagcccccattacacagtgagctgtgagcttccaaaacacgcgtgagagattgc 1213
Oy 1012 ttctccctagacatcgtagtgaatgatacctggtgtgtgtgtgtgtgtgtgtgtgtgt 1071
Db 1214 ttcgcatcgaatggtgtgactgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1273
Oy 1072 aagcacggtctcgaagctgtagatcgcacacagagatgactgctcgtctgcacacatc 1131
Db 1274 aaatacgtatctcaactgtagtgcacatcagatagatagactgtgtgtgtgtgtgtgtgt 1333
Oy 1132 ttgtgtctctcttcaaatcatgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1191
Db 1334 gtctgtgtgagatctcagagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1393
Oy 1192 gatgtgagctgtagaagaataatccaggtgtgcagcgtgtgtgtgtgtgtgtgtgtgtgt 1251
Db 1394 gagaacacagagagcaaaacacagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1453
Oy 1252 accatgtcgtctcgtgagatctatctgtatctctcctcaagctgtgtgtgtgtgtgtgt 1311
Db 1454 gtgtgtctgagcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1513
Oy 1312 atcgtgtcaatctcaagaactccctcaagaactcaccgacccctactacctgtgtgag 1371
Db 1514 gcatgtggaacttaagaggaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1573
Oy 1372 aagagcagcgtgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1431
Db 1574 aagagcaaatatgtattgttaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1633
Oy 1432 ctgcgcctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1491
Db 1634 ctgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1693
Oy 1492 cagttcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1551
Db 1694 caatttccaaatgtgagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1753
Oy 1552 aagacctaatataggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1611

```

```

Db 1754 aaagattattatgatalgtatgagccagaagagtgaaatttcaagatgtccatctcct 1813
Oy 1612 ctctacttgcgaactcagagatcttcagggcaaaaggtcatcg 1654
Db 1814 atctacttgcgaacatgtgttcttcttagcggaactatcg 1856

RESULT 5
AAX78074
ID AAX78074 standard; cDNA; 2792 BP.
XX
AC AAX78074;
XX
DT 19-AUG-1999 (first entry)
XX
DE Rat DTDST cDNA.
XX
KW DTDST; human, expression vector; sulphate transporter; screening;
KW bone disease; cartilage disease; treatment; cell sulphate ion intake;
KW drug preparation; rat; ds.
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 368..2587
FT FT /*tag= a
FT FT /product= "DTDST"
XX
PA JP11146790-A.
XX
PD 02-JUN-1999.
XX
PF 18-NOV-1997; 97JP-0335157.
XX
PR 18-NOV-1997; 97JP-0335157.
XX
PA (SUMU ) SUMITOMO SEIYAKU KK.
XX
DR WP1; 1999-378999/32.
XX
DR P-PSDB; AAY08929.
XX
PT Sulfate transporter gene expression vector
XX
PS Example 1; Page 16-19; 22pp; Japanese.
XX
CC This invention describes the construction of a novel vector for sulfate
CC transporter expression containing a DNA sequence encoding a mammalian
CC sulfate transporter (expression product of the DTDST gene) and
CC containing no DNA sequence of 5' translation region of the mammalian
CC sulfate transporter gene. The invention also describes: (A) a method for
CC screening a human bone/cartilage disease treating agent including the
CC steps: (1) transforming an animal cell with the above vector, (2)
CC culturing the animal cell in the presence of a sample and (3) detecting
CC the increase in the sulfate ion intake to the cell; (B) a drug
CC preparation for the treatment of human bone/cartilage disease containing
CC the above vector as the active component. The sulfate transporter
CC gene-containing vector is high in expression efficiency. This sequence
CC encodes the rat DTDST protein used in the method of the invention.
XX
SQ Sequence 2792 BP; 694 A; 651 C; 661 G; 786 T; 0 other:

Query Match 10.7%; Score 242.2; DB 20; Length 2792;
Best Local Similarity 49.3%; Pred. No. 2e-50;
Matches 775; Conservative 0; Mismatches 758; Indels 38; Gaps 4;

Oy 106 aaacttgcgaatgcttcaagatgttctcctcaagcgaagataaagctgtgtgtgtgtgtgtgt 165
Db 584 aagcttcagaagagtggtccagtgtaatgcaacaaatcagaatagatgtttttgattt 643
Oy 166 ctgctgtgtctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 225
Db 644 ttctctgtttgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 703

```

QY 226 ctcggtggaactcagcgagggaatccatccaggtccccaagaagcattgctcgcg 285  
 Db 704 atgtctgcgcgtatgttggttatattgtgtgcccagtcattcttaaccctgtg 763  
 QY 286 gccaaactctcgcagtcataatgctctactcctctctccctcctgactacttc 345  
 Db 764 gctgcgaagaacctatctatgtctgtaacatacttlttgccagcatattacttc 823  
 QY 346 ttccctgggggtgttcacagaaatggtccaggtactttgcgttataccagctcggtg 405  
 Db 824 cgtgttgcatacctccgcacacatctctgtggaattttggaactgtgcttaagatc 883  
 QY 406 ggt-----aacatctgtcagctggtcccgagagtcgaaatcca-----g 448  
 Db 884 ggtgaagtagttgacccgagaaactaataagcctgcccgcgaattgatactaatatc 943  
 QY 449 tcttcaacaatgcaccaatagagctatgt-----ggaacagc 488  
 Db 944 tccatagcaattgttccaatgtgatlgtgtagtaaacatatacatataagcgaactctgt 1003  
 QY 489 agccatggaagcctgagaagcctgacgtgtcagctagcctgctcctcaaccctatc 548  
 Db 1004 gacaaaagctgttatgcaattaaatggaagc-actgtgacatcattaggtgtgagttta 1062  
 QY 549 ccagatggtgtcgtggtctcattgcaatgtgtgtgtggtcattcctcagatcctt 608  
 Db 1063 tcagtagcagcattggtgtcttctcaagtgtggtctgtgtctgtactcctcagatgctt 1122  
 QY 609 catccgggtctcattgagcgccggcctgtgcagatccctgatttgcgtcagatcat 668  
 Db 1123 gctgaagcgggttgtgtactgtgtcctcctcaacatccctcaagtcctcagcgatccct 1182  
 QY 669 ctctgagcagcagcatccctcctcctacacagcccaaggtccatcgtcttactccttga 728  
 Db 1183 cctggggtgcgagcctcctcccgagcaatgtgtgagctcagtcattactacatcggttcca 1242  
 QY 729 catttgcanaaacctccccaacacaaactcgcctcgtcactcctgtctcctcagcg 788  
 Db 1243 catcttcagaataatcatataagacacaacatctgtgacccatcaccagccttltgtct 1302  
 QY 789 tgccttccctggtgtgtgtgaagagatcattgtcgtctacatgacaaagatctgctcc 848  
 Db 1303 cctgttcccttgtgcacaacaaagaactcaacgaatacttccaagtccctccgagcaco 1362  
 QY 849 catccctacagagatgt 908  
 Db 1363 aattccaactggtcattgt 1422  
 QY 909 gcccaaaaagtacacatgcagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 968  
 Db 1423 caattggaattgaattcagatattgtcgggcgaatactccacacgggtttatgaccacca 1482  
 QY 969 gtcgcctgt 1028  
 Db 1483 agcgcagagcgtgagcctcaatctcctaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1542  
 QY 1029 gagctacgtcatcaactgt 1088  
 Db 1543 tgttttgttatcacgtatacttctgtgagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1602  
 QY 1089 ggaattcgaacagagatgtatcgtctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1148  
 Db 1603 gaaagcaactcagaagtgtatgtccattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1662  
 QY 1149 aattcatgtcattgt 1208  
 Db 1663 ctgcataactatgt 1722  
 QY 1209 atcccaagtgccagcagctgt 1268  
 Db 1723 gaccacaactgtcgcgtatagtgaatccctcgttcttctgtgtgtgtgtgtgtgtgtgtgtgt 1782

QY 1269 gatctatctgtatccctcctccttaagtcgtgtcaggagccctgatacgtgtlcaatctcaa 1328  
 Db 1783 tcccttatctatctccctccaaaagtgtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1842  
 QY 1329 gaactccctcaagcaactcaaccgaaccttaacctgtgtgtgtgtgtgtgtgtgtgtgtgt 1388  
 Db 1843 ggtgtgactctgtgaaatttagaagacctgtccaaagatgtgtgtgtgtgtgtgtgtgtgt 1902  
 QY 1389 ttgcagctgt 1448  
 Db 1903 agttatctgt 1962  
 QY 1449 agt 1508  
 Db 1963 cgttgggt 2022  
 QY 1509 tgcactggccaggt 1568  
 Db 2023 actgctgt 2082  
 QY 1569 ccaggaatccagggattaaatcaatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1628  
 Db 2083 tgggaatgaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2142  
 QY 1629 agagatcttca 1639  
 Db 2143 agaatgtctta 2153

## RESULT 6

AA11766 standard; DNA; 1465 bp.

AA11766;

12-OCT-2001 (first entry)

Probe #1699 for gene expression analysis in human cervical cell sample.

Probe: human; microarray; gene expression; cervical epithelial cell;

cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

09-AUG-2001.

30-JAN-2001: 2001WO-US00670.

04-FEB-2000: 2000US-0180312.

26-MAY-2000: 2000US-0207456.

30-JUN-2000: 2000US-0608408.

03-AUG-2000: 2000US-0632366.

21-SEP-2000: 2000US-0234687.

27-SEP-2000: 2000US-0236359.

04-OCT-2000: 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID NO 1699; 487bp; English.

The present invention relates to human single exon nucleic acid probes

(SENP). The present sequence is one such probe. The SENPs are derived

from human HeLa cells. The SENPs can be used to produce a single exon

microarray, which can be used for measuring human gene expression in a





Matches 556; Conservative 0; Mismatches 599; Indels 0; Gaps 0;

```

QY 546 catcagatgggtcgtggtctcagttcagttgtggtcgtcaggtcaggtc 605
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 111 cttccaggtagcgatgggtctcttcaagtgggtttgtctgtctactccagatgc 170
QY 606 cttcaatccggggctcattcagcgccggcctgcagactcgtattcgtgtcgaagta 665
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 171 ctgctgagtgatgttgcactgtgtcctcctcactatcttaactcagccagta 230
QY 666 catcttcgagtagcacatccctccctcctcagagccaggttcattcttactcat 725
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 231 tctctctgggtcctaacccttcctcggaactaagtgtgtggtcctcactcactcagat 290
QY 726 tgacatttgcaaaaaccccccacccaacatcgctcgtcattcgtctcactcag 785
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 acattgtcttcagaaacatccatagacaaatctctgtatctatccagagcctttg 350
QY 786 cgtgtcctcctcgtgtcgtgtgaagagctcaatgtcgtctacatgcacaagctcgctt 845
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 351 cctttgtgtcttcttgccaacaaagaaactcaatgaacacttcaatccagcttaagc 410
QY 846 ccccatccctacagagatgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 905
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 411 accgattccatgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 470
QY 906 gatgccaaaagtaacacatgcagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 965
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 471 actacaatgaataatataatcttagtatgtctgtgacatattccacgtgtgtgtgtgt 530
QY 966 ggtgtcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1025
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 531 caaagtaacagaatgtgaactaatctcagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 590
QY 1026 cgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1085
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 591 catgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 650
QY 1086 cgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1145
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 651 agtcaagaacaaacagaatgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 710
QY 1146 taataatcattgtcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1205
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 711 ccaatttctactactagtagcagctctgtgaagacattgtgttaagaatcaacaagctgt 770
QY 1206 aaatccaggtgtgcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1265
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 771 ccatctacagcttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 830
QY 1266 ggggagatctatctgtatcctctccctaagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1325
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 831 agctccttctgtctatctccttcaaaaaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 890
QY 1326 caagaactccctcaagaactcaacccctcactacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1385
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 891 acggggagccctcgttaaatlttagggatcttcccaaaagtgtagatattagtagaattga 950
QY 1386 ctgtgtcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1445
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 951 taacagtattcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1010
QY 1446 ggcagatgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1505
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1011 actgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1070
QY 1506 ctatgtacatggccaggtatgcacatgtacattatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1565
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1071 ttcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1130
QY 1566 ggcacaggtatccagggatgaataatcatcagttactgtcctccctctactgttgcga 1625
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1131 ccttcagactaagcagagcatcaagattcttcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1190

```

QY 1626 ctcaagatcttcaggaacaaaggtcagccagagctgtcctcgtcaggaagctgcagca 1685  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 1191 caaagatgtctttaaactcgtgttatacaaaacaaactgtcaacccaacttaataaaggt 1250

QY 1686 ggaacttggagaatgc 1700  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 1251 ggccttggaaagaagc 1265

RESULT 8  
 AA101695  
 ID AA101695 standard; DNA; 1465 BP.  
 XX  
 AC AA101695;  
 XX  
 DT 09-OCT-2001 (first entry)  
 XX  
 DE Probe #1686 used to measure gene expression in human breast sample.  
 XX  
 KW Probe; human; breast disease; breast cancer; development disorder; ss;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157270-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US000661.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression  
 XX  
 PS in a human breast -  
 XX  
 PS Claim 25; SEQ ID No 1686; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast.  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX  
 XX  
 SO Sequence 1465 BP; 389 A; 319 C; 284 G; 473 T; 0 other;

Query Match 8.7%; Score 196.6; DB 22; Length 1465;  
 Best Local Similarity 48.1%; Pred. No. 3.5e-39;  
 Matches 556; Conservative 0; Mismatches 599; Indels 0; Gaps 0;

QY 546 catcagatgggtcgtggtctcagttcagttgtggtcgtcaggtcaggtc 605  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 111 ctccaggtgagatgaggtctcttcaagtggttcttctcttcaactcctcaagatgc 170  
 QY 606 ctccatccggggcttcataagcgcgcgccttcagatccatcttctggtcgaagta 665  
 Db 171 ctgctgagtgaggttctgcatctggtcctctcttcaactatcttcaatccagcgaagta 230  
 QY 666 catctctgagatgacatccctccctcaacagccagggatccatcgcttcaactcat 725  
 Db 231 tctctctggtctcaactctcctcgactaatgtgtgtggtcctcaactcatcaactgtgat 290  
 QY 726 tgacattgcaaaaacctcccccaacaatctgcctcgctctcttctgcttcatcag 785  
 Db 291 acatgtctcaagaacaatccataagaccaatctctgtgactcttaccagccttctgtg 350  
 QY 786 cgtgtcctctctgtgtgtgtgaagagctcaatcgtctgctacatgccaagatctgtc 845  
 Db 351 ccttctgttctcttctgccaaccaagaactcaatgaacacttcaatccaaagcttaagc 410  
 QY 846 ccccatccctcaagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 905  
 Db 411 accgatccctcttgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 470  
 QY 906 gatgccccaaaagatatacatatgcagatcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 965  
 Db 471 actatcgaataaataatataatctagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 530  
 QY 966 ggtgtcgcctgt 1025  
 Db 531 caaagatcccgaaatgaaacctaattcctctagtggtgtgtgtgtgtgtgtgtgtgtgt 590  
 QY 1026 cgt 1085  
 Db 591 catgt 650  
 QY 1086 cgt 1145  
 Db 651 agtcaaaagcaaaacgaagaatgtatgcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 710  
 QY 1146 taaatcatgtcatctgt 1205  
 Db 711 ccaactgtttactactagt 770  
 QY 1206 aaaaatcccaagtggt 1265  
 Db 771 ccatctcagcttctctgt 830  
 QY 1266 ggggatcatctgtatctctccctcaagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1325  
 Db 831 agctccttctgtctatctcccttcaaaaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 890  
 QY 1326 caaagaactccctcaagaacactcacaccactactactctgtgtgtgtgtgtgtgtgtgtgt 1385  
 Db 891 accgggggacctctgttaatttaggtatcttcccaaatgtgtgtgtgtgtgtgtgtgtgtgt 950  
 QY 1386 ctgtgtcatctgt 1445  
 Db 951 tacagttatctcgt 1010  
 QY 1446 ggcagtggt 1505  
 Db 1011 actgt 1070  
 QY 1506 ctatgactgt 1565  
 Db 1071 ttaactgt 1130  
 QY 1566 ggcacagagatatacagggatataaatacatcagtaactgtccctctctactgtgtgtgt 1625  
 Db 1131 ccttcagactaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1190  
 QY 1626 ctgcagagtgctgt 1685  
 Db 1191 caaagaatgtcttataatctgtcttatacaaaaacagtgtaaccaacatcttaataaaggt 1250

QY 1686 ggaacttgagaatgc 1700  
 Db 1251 ggccttgagaagaagc 1265  
 RESULT 9  
 AAI20978  
 ID AAI20978 standard; DNA; 1521 BP.  
 AC AAI20978;  
 DT 12-OCT-2001 (first entry)  
 DE Probe #10911 for gene expression analysis in human cervical cell sample.  
 DE XX  
 KW Probe: human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.  
 OS Homo sapiens.  
 PN MO200157278-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00670.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 XX  
 PS Claim 25; SEQ ID No 10911; 487bp; English.  
 CC  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SNP). The present sequence is one such probe. The SNPs are derived  
 CC from human HeLa cells. The SNPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 1521 BP: 412 A; 325 C; 309 G; 475 T; 0 other;  
 XX  
 Query Match 8.5%; Score 192.2; DB 22; Length 1521;  
 Best Local Similarity 48.0%; Pred. No. 4.4e-38;  
 Matches 548; Conservative 0; Mismatches 593; Indels 0; Gaps 0;  
 QY 560 tgggttcaagcaggt 619  
 Db 8 tgggttcttcaagaagtggt 67  
 QY 620 tcatcagcgccgc 679  
 Db 68 ttgtacacgt 127  
 QY 680 ccatccctctcaacagggccagggatccatcgttcttaactcttcaattgtgcaaaa 739

Db	128	acctctctcgactcaatcggtggtggtctcaactcaactcaactccggatacatgtctctcaaa	187
Oy	740	acctccccacaacaacatcgccctcgctcaactctctgctctcaacagcggtgctctccgg	799
Db	188	acatccataagaaccaatctctgtgactcttatacccaagccctttgtgcttctgtctctt	247
Oy	800	tgctggtgaaagagctcaatgtctcgtctacatgcaagaatcgtctcccaatccctacag	859
Db	248	tgcacaaccaagaagactcaatggaacactctcaaacaccgaagcttaaggacacggaattccattg	307
Oy	860	agatgatgtgtgtgtgtgtgtgcacaaagctcaatcccggtgcgtgtgaagtcccaaaagt	919
Db	308	aacttgtgtgtgtgtgtgtgcacaaagctcaatcccggtgtgtgtgaagtcccaaaagt	367
Oy	920	atcacatgacagatcgtgtgtgtgtgtgtgcacaaagctcaatcccggtgtgtgtgaagt	979
Db	368	ataatctcagatgtgtgtgtgtgtgcacaaagctcaatcccggtgtgtgtgaagt	427
Oy	980	tctcacagtggaagagacatgatatgacacagcctctctccatgacatcgtgtgactagtica	1039
Db	428	ggaacactcaatctcctgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	487
Oy	1040	tcaaccttgagctatgt	1099
Db	488	tcaactgtatcaactctctcgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	547
Oy	1100	aggagatgatcgtctcctcggt	1159
Db	548	aggaaatgtatgcatctgt	607
Oy	1160	tttgt	1219
Db	608	ctagtgtgcagctctctgt	667
Oy	1220	ccagacgt	1279
Db	668	ctgt	727
Oy	1280	atctctctccaaagtcgt	1339
Db	728	atctctctccaaagtcgt	787
Oy	1340	agcaactcaacgcagccctactaccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1399
Db	788	gtaaatatgagatctctcccaaaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	847
Oy	1400	tagtgagct	1459
Db	848	tgtgtactatgtct	907
Oy	1460	ct	1519
Db	908	gttttctctatatttgt	967
Oy	1520	aggtcatgagacatgacattatgt	1579
Db	968	tgt	1027
Oy	1580	aggggatataaatacatcagtaactgtctccctctctctctctctctctctctctctctctct	1639
Db	1028	caggacataaagatct	1087
Oy	1640	ggcaaaaggtcatcgcgaagagctgtctctctctctctctctctctctctctctctctctct	1699
Db	1088	aatctgtcttatatacaacaacatcgtaaccaactcttaataaagtggtgtgtgaagaag	1147
Oy	1700	c 1700	
Db	1148	c 1148	
RESULT	10		

[illegible]



```

Db      368  ataattcagtaattcgtcgacacatccaccatcgggtttaaagccaccacaaagtaccagaat 427
Oy      980  tctcaagtgaggaagacagatagagacagaccttccctccagcagcagcgtgaatcagta 1039
Db      428  ggaacctaatcttcctgctgctgctgctagatgacatattccacatattgcttctgta 487
Oy      1040 tcaacctgtgctatggcgccgaccttggccaaagacagcagctgacgctggtgaattcgacc 1099
Db      488  tcactgtatcacttcttcgtgagatgtttggccaagaacaatggttaacagtcgaagcaaac 547
Oy      1100 aggaatagtcgctcctcgtcgcagcaactcttggcctctttaaattcaatcatgtca 1159
Db      548  aggaatgtatgcattcgtccttcttcaatataatcccttccctcttccactgttttacta 607
Oy      1160 ttgtcgtgtcgtcttcctcgtcagcctgctgctgagtgagtgagtgaggaataatcccaagg 1219
Db      608  ctatgtagcagctcttgcgaagacatgtgttaagaatcaacaggctgcactactagctt 667
Oy      1220 ccaagctgtgtgtgtcctcgtgctgagatgataccatgctgtcctcctggagatcctgt 1279
Db      668  ctggtgtgtgtaacagccctgcttcttctgttctcctcctagtaatagctcttcttct 727
Oy      1280 atcctcctcctaagctcgtgctgctgagagccctgctgctgctcaatctcaagaactccctca 1339
Db      728  atcccttcaaaaaggtgctccttggctgtgatacacaattgtaaatctacggtgagcccttc 787
Oy      1340 agcaactcacgaccctactacccctgctgaggaagagagcaagcttgagctgtgacatgg 1399
Db      788  gtaaatcttaggagatcttcccaaaatgtgtagatagtaagaatgatatagctatctggt 847
Oy      1400 tagtagactctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1459
Db      848  ttgttactatgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 907
Oy      1460 ccttctcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1519
Db      908  gtttctctatatttgtgtcattcctcctcctcctcctcctcctcctcctcctcctcctcct 967
Oy      1520 aggtcatgacactgacattatgtgtaatcccaagaccataatgtggcgccggatattcc 1579
Db      968  tggtaggaagagctcgtgagcttcttgaaatcgtgtcgttacaagaactcctcagactaagc 1027
Oy      1580 agggagattaaatcatcactgactgtccctcctcctcctcctcctcctcctcctcctcctc 1639
Db      1028 caggaatcaagatttctcgtcttgtagccctcctcctcctcctcctcctcctcctcctc 1087
Oy      1640 ggcataaaggtcactgcgaagactgtcctcctcctcctcctcctcctcctcctcctcctc 1699
Db      1088 aatcgtcttatacaaaaactgtcaaccacacttaataaagctgtgtaagaagg 1147
Oy      1700 c 1700
Db      1148 c 1148

```

RESULT 12  
AA159022  
ID AA159022 standard; cDNA; 1466 BP.

AC AA159022;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1225.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.

```

OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0596042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HNS-E) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Qa, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPT, 2001-442253/47.
XX
PT P-PSDB; AAM39866.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 1225; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S. disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
SQ Sequence 1466 BP; 367 A; 399 C; 389 G; 310 T; 1 other:

```

Query Match 5.4%; Score 122.6; DB 22; Length 1466;  
Best Local Similarity 58.7%; Pred. No. 9.4e-21;  
Matches 212; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```

Oy      1280 atcctcctcctaagctcgtgctgtagagccctgtagctgtcaatctcaagaactccctca 1339
Db      11  acccaagcgtccggcggtcctcgtgacacatcatcatctgtgactgaaggcagctgca 70
Oy      1340 agcaactcacgaccctactacctcgtgtgaggaagcaagctgtagctgtgacatggg 1399
Db      71  ggcagctcagcgacatgctcctcctccttggaaagccaatcgggcgatcgtcttacttggc 130
Oy      1400 tagtagactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1459
Db      131  tggtagactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 190
Oy      1460 ccttctcgtcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1519
Db      191  tcttctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 250
Oy      1520 aggtcatgacactgtatctgtaatcccaagaactataataggtcccgagatattcc 1579

```

Db	251	aggtgcccagaccgagatatttcacagagatgtgacagatcactcaaggccaaaggaagtc	310
Qy	1580	aggggattcaaatcatcacgtactgtccctctctactcttggccaactcagagatcttca	1639
Db	311	gggggggtgaaggtcttcgcctctccgcgcacacggtgacttggcaatgctgagttctaca	370
Qy	1640	g 1640	
Db	371	g 371	
RESULT 13			
ID	AAI60808/c	AAI60808 standard: cDNA, 1227 BP.	
AC	AAI60808;		
XX	22-OCT-2001	(first entry)	
DE	Human polynucleotide SEQ ID NO 4797.		
XX	Human; nocotropic; immunosuppressant; cytosstatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokineic; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukaemia; ss.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
FN	26-JUL-2001.		
PD	26-DEC-2000; 2000MO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
PA	(HXSE-) HXSEQ INC.		
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodfich R, Drmanac RT;		
XX	WPI: 2001-442253/47.		
DR	P-PSDB; AAM41652.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX	Claim 1; SEQ ID NO 4797; 10078pp; English.		
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nocotropic,		
CC	immunosuppressant and cytosstatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: immune system suppression.		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		

CC	Note:	The sequence data for this patent did not form part of the printed specification.
CC		
XX		
XX	Sequence	1227 BP; 254 A; 342 C; 350 G; 281 T; 0 other;
SQ		
	Query Match	5.4%; Score 122; DB 22; Length 1227;
	Best Local Similarity	59.5%; Pred. No. 1.2e-20;
	Matches 206; Conservative	0; Mismatches 140; Indels 0; Gaps 0;
OY	1295	cctgctcgtaggaaccccttgatcgctgtcaactctaagaacctcccctaagaacctcacaccgacc 1354
DB	1226	CGCTCCTGGCACCATTCATCATTTGTGAACCTGAAGGCGATGCTGAGAGCGCTTAGCGACA 1167
OY	1355	cctactaacctctggaggaaagacaagtctgactgttgcatacttgggttagtgaatttcctct 1414
DB	1166	TGCGCTCCCTCTTGGAAGGCCAATCGGGGGGACTGCTTATCTGAGCGTGGTGAACCTTCACGG 1107
OY	1415	cctcctctctccctcagacctgacctgatgtgtgtgacagtgggtgtgcctttccctccgtcgtg 1474
DB	1106	CCACATCATCTGCTGAAACCTGAGACCTTGCTGGTGGTGCCTGCGGCATCTCTCCCTGCTGC 1047
OY	1475	tctgtgtcttcacagactcagctcagtttcgaaaatggcctatgcacatgcctgccccaggtcatgagacatg 1534
DB	1046	TGTGTGTGTGTGGGACACAGATGCCCCCACTACTCTGTCTCTGGGGCAGGTGCCAGACACGG 987
OY	1535	acatttatgtgataccccaagacctataatagaagccccagatatccagaggatlaaatca 1594
DB	986	AATATTTCAGAGAGATGTGGGACAGTACTCAGAGGCGCAAGAAGTCGGGGGGTGAAGTCT 927
OY	1595	tcaagtaactgtctccctctctactattgtccaactcagagatcctcag 1640
DB	926	TCCGCTCCTCGGGCCACCCTGTACTTTCGCAATGCTGAGTTCTACAG 881
	RESULT 14	
AAAI6124	ID	AAAI6124 standard; DNA; 588 BP.
XX	AC	AAAI6124;
XX	XX	
DT	14-JUN-2000	(first entry)
XX		
XX	Human colon cancer differentially expressed nucleotide sequence #129.	
DE		
XX	Colon cancer: detect; differential expression; human; treatment;	
KW	detect mutation; non-invasive diagnostic method; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200012702-A2.	
XX		
PD	09-MAR-2000.	
XX		
PF	30-AUG-1999; 99WO-US19424.	
XX		
PR	31-JAN-1998; 98US-0098639.	
XX	27-JAN-1999; 99US-0117393.	
PA	(FARB ) BAYER CORP.	
XX		
PI	Endege WO, Steilmann KE, Astle JH, Burgess CC, Carroll E;	
PI	Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;	
PI	Schlegel R.	
XX		
DR	WPI; 2000-256641/22.	
XX		
PT	Novel nucleic acids and proteins for identifying therapeutic agents	
PT	useful for treating and diagnosing cancer, especially colon cancer -	
XX		
PS	Claim 16; Page 176; 345pp; English.	
XX		
CC	This sequence represents a human nucleotide sequence which is	





! |||  
Db 482 actatcg 489

Search completed: April 25, 2002, 14:41:58  
Job time: 6000 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:32:31 ; Search time 3196.3 Seconds

(without alignments)  
11674.943 Million cell updates/sec

Title: us-09-749-589-1

Perfect score: 2262

Sequence: 1 atgagccagccagccagccgcg.....agaccctgacgcgcctgtga 2262

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_ov:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rod:\*  
33: em\_htgo\_hum:\*  
34: em\_htgo\_inv:\*  
35: em\_htgo\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	392.4	17.3	2855	10	AF337809	AF337809 Rattus no
2	385.6	17.0	2589	9	AF288410	AF288410 Homo sapi
3	385.6	17.0	2748	9	AF279265	AF279265 Homo sapi
4	364.4	16.1	2629	10	AF136751	AF136751 Mus muscu
5	358	15.8	4113	10	AF230376	AF230376 Meriones
6	355.8	15.7	2612	10	AY032863	AY032863 Mus muscu
7	343.4	15.2	2853	10	AF167412	AF167412 Rattus no
8	339.4	15.0	3086	10	AF167411	AF167411 Mus muscu
9	337.2	14.9	2235	10	AY024359	AY024359 Mus muscu
10	334.6	14.8	1389	4	AF314819	AF314819 Oryctolag
11	332.8	14.7	2235	10	RMO303372	AF303372 Rattus no
12	332.4	14.7	1634	10	AF314820	AF314820 Rattus no
13	328.2	14.5	4930	9	AF030880	AF030880 Homo sapi
14	326.2	14.4	2881	9	HUMDRA	L02785 Homo sapien
15	326.2	14.4	2881	11	G18336	G18336 swiss51883 Er
16	326.2	14.4	2882	6	AR052312	AR052312 Sequence
17	326.2	14.4	2882	6	AR144265	AR144265 Sequence
18	326.2	14.4	2882	6	I28252	I28252 Sequence 1
19	289	12.8	62357	2	AL451008	AL451008 Homo sapi
20	289	12.8	62377	2	AL360009	AL360009 Homo sapi
21	259.6	11.5	2205	4	OAR18558	Y18558 Ovis aries
22	242.2	10.7	2792	6	E26555	E26555 DTDST gene
23	242.2	10.7	2877	10	D82883	D82883 Rattus norv
24	230.2	10.2	3076	9	AF297659	AF297659 Homo sapi
25	225.8	10.0	4741	4	B07223615	AJ223615 Bos tauru
26	223	9.9	3711	10	RATSAT1A	L23413 Rattus norv
27	222.8	9.8	3732	10	MUSSTOB	DA2049 Mouse mRNA
28	213	9.4	2832	9	HSU14528	U14528 Human sulfa
29	211.2	9.3	1259	10	AF315652	AF315652 Rattus no
30	204.4	9.0	178189	2	AC019103	AC019103 Homo sapi
31	196.6	8.7	175550	9	AC008427	AC008427 Homo sapi
32	196.6	8.7	235808	2	AC011406	AC011406 Homo sapi
33	190.2	8.4	165443	2	AC079939	AC079939 Mus muscu
34	182.8	8.1	185081	2	AC079937	AC079937 Mus muscu
35	143.2	6.3	185081	2	AC079937	AC079937 Mus muscu
36	119.4	5.3	32037	3	AC018283	AC018283 Drosophil
37	109.8	4.9	40197	1	CBRG33P21	AE004592 Caenorhab
38	98.4	4.4	10208	1	AE004592	AE004592 Pseudomon
39	85.6	3.8	3008	9	HSMB00532	AI050358 Homo sapi
40	76.6	3.4	39874	3	CEPK287	Z70757 Caenorhabdi
41	73.8	3.3	2403	9	AK023366	AK023366 Homo sapi
42	71.6	3.2	2515	8	SSSULPTRP	X96761 S.stapifianu
43	71.4	3.2	88326	9	AC005923	AC005923 Homo sapi
44	71.4	3.2	204513	2	AC024179	AC024179 Homo sapi
45	71	3.1	89203	2	AC010691	AC010691 Drosophil

## ALIGNMENTS

RESULT 1  
LOCUS AF337809 2855 bp mRNA  
DEFINITION Rattus norvegicus down-regulated in adenoma protein mRNA, complete cds.  
ACCESSION AF337809  
VERSION AF337809.1 GI:15077506  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2855)  
Ye,H.J., Binder,H.J. and Rajendran,V.M.  
AUTHORS Molecular cloning and characterization of down-regulated in adenoma (DRA) mRNA from rat colon  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2855)  
Ye,H.J., Binder,H.J. and Rajendran,V.M.  
AUTHORS







Db	1368	CAGTGCCTTCCCCGATGTTGCTTATGTCTCGGAGACCTGGTACAGGAGAAGACCGGGGGC	1427
Oy	1207	aaatcccaagtgagccagccctgtgtgtctctgtgtgtgtatcatcaccatcgtgtctcgt	1266
Db	1428	AACTCGAGGTTGGTGGAGGCCATCTCTCCCTTTTCATCTCCCTCATCATGTGTAAACCTT	1487
Oy	1267	gggatactctcgtataccctccctccctaagctctgtgtcagaagaccctgatacgtcgtcaatc	1326
Db	1488	GGGGAACTCTTCCATGACACTGCCCCAAGGCGGCTCGGAGCCATCATCATGTGTGAACCTG	1547
Oy	1337	aagaactcccccacaactcaaccgaccctactaccctcgtgtgaggaagaagcaagcttgac	1386
Db	1548	AAGGCGATGCTGTAGGACACTACGACATGGCGCTCCCTCTGTGAAGGCCAATCGGGCGGAT	1607
Oy	1387	tgttgacatccttggtatgtagtgaattcctcctcctcctcctcctcctcctcctcctcctc	1446
Db	1608	CTGCTTATCTGCGCTGGTGAACCTTACGGCGACCATCTTGCTGACCTGAGACCTTGCGCTTG	1667
Oy	1447	gcagtgaggtgtgtgcctcttcctcgtctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1506
Db	1668	GTGGTGGGCGTCATCTTCTCCCTGCTGCGGGGCGGTCGGACACAGATGCCCCACATAC	1727
Oy	1507	tatgacatcggcccagatcatatgacactgacatattatgtgaatcccaagcattataatag	1566
Db	1728	TCTGTCTTGGGGCAGGTGCCAGACACGGATATTTCACAGAGATGTGGCAGATCTACAG	1787
Oy	1567	gcccagatatccaggaggatlaaatalcatcaagtaactgtccctcctcctcctcctcctc	1626
Db	1788	GCCAAAGGAAGTCGGGGGGGTGAAGGTCTTCCGCTCTCTGGCCACCGTGTACTTGGCAAT	1847
Oy	1627	tcaagatccttcag	1640
Db	1848	GCTGAGTCTACAG	1861
RESULT 4			
AF136751			
LOCUS	AF136751	2629 bp	02-AUG-1999
DEFINITION		Mus musculus down-regulated in adenoma protein (Dra) mRNA, complete	
ACCESSION	AF136751		
VERSION	AF136751.1	GI:5359729	
KEYWORDS			
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 2629)	
TITLE		Melvin,J.E., Park,K., Richardson,L., Schultheis,P.J. and Shull,G.E.	
JOURNAL		Mus down-regulated in adenoma (DRA) is an intestinal	
MEDLINE		Cl(-)/HCO(3)(-) exchanger and is up-regulated in colon of mice	
REFERENCE		lacking the NHE3 Na(+)/H(+) exchanger	
AUTHORS		J. Biol. Chem. 274 (32), 22855-22861 (1999)	
TITLE		2 (bases 1 to 2629)	
JOURNAL		Melvin,J.E., Park,K., Richardson,L.A., Schultheis,P.J. and	
FEATURES		Shull,G.E.	
SOURCE		Direct Submission	
		Submitted (23-MAR-1999) Center for Oral Biology, University of	
		Rochester, 601 Elmwood Avenue, Rochester, NY 14642, USA	
		Location/Qualifiers	
		1..2629	
		/organism="Mus musculus"	
		/strain="FVB/N"	
		/db_xref="taxon:10090"	
		/sex="female"	
		/tissue_type="colon"	
		/dev_stage="adult"	
		1..2629	
		/gene="Dra"	
		131..2404	
		/gene="Dra"	
CDS			

Query Match	Best Local Similarity	Score	DB	Length
Matches 863; Conservative	52.8%;	364.4;	DB 10;	2629;
	0;	Mismatches 761;	Indels 12;	Gaps 3;
22	tacgtggtatagacagcgccgacacaccccttaccctctcgacagatgattgagagaag	81		
155	TTATGTGTGGTGGCCAGACCGGTGTACTCCACAAAGACTTTTGGAGAGAGATTAAAGACG	214		
82	gaccgacatcacccagttggagagaacttcgaatg---ccctaaatgctccagcc	138		
215	CACACACATACACAGACATTTCTCTGGACCATCTCCAAAGGGCTTTGTAGCTGCCTCCACAG	274		
139	aagatcaaaagcgtgtgttttggtgctgcgcctgtgcctctctgctccccaagatagaag	198		
275	AAGGCCAAGAAAATGCGCCTGTCTTTGTTCCCATAGCATCTTGGTTGGCAGCCTACAAAG	334		
199	atlaaagacatcatctctcctgacctgcctgcgttggaactagagcggggagatccatccaagtc	258		
335	ATTAAGAGATGGCTTCTCAGTGACATCGTCCTGGCATCAGCAGCCGGGTGGCTGTGA	394		
259	ccacaagacagtgcatctgtcctgcctgcgcacacctctcctgaagtaatgacctactcc	318		
395	CTGCAAGGTTTAGCATTTTGTCTGCTGTCGTCACATCCCTCCAGCCGTACGGGTTGTACGCA	454		
319	tccctctcccccctctgacctactctctcccgaggggggtctcacagatggtgccaggt	378		
455	GCCTTTTCCTCCAGTATTAACCTACTTTTCTTGGGCACATCTAGACACTATCTGTGGGT	514		
379	acctctgcgctatcagaactcctgcttggttaacatctgtcctgcagctgagcccaagatcg	438		
515	CCATTTTCCAGTTCTGTGATATGATGGGGAGTTGTCTGTTTCAAGAGATGGTCTCGGACCCC	574		
439	aaattccagctctcaacaatgcacccaatgagagctatgtgagacacagcagccttgag	498		
575	AATGCT----TCTTCAGAGTTGTCTCCA-----AGCTCCACAGAAATGACTCGTTCAATA	625		
499	gcttgagagctgacagtgctcagctacgctagcctgacctcaaccgcatcatccagatgggt	558		
626	GAGGAGAAAGTATGTGTGCTGCATACATGATCTTTCTTGTGATATCATTCATGATTGGTCC	685		
559	ctggagctcatcagtttggctcttggtgcacatccctcctcagatctctcaatccggggc	618		
686	CTGGGGGGTTCTCAGTGGGCTTTGGTGGATATACCTATTCGAGAGTCCCTATATCAGTGGC	745		
619	ttcatacagcgccgagccttgacagatccatgattcctggtcctcaagtatcatcttcgagct	678		
746	TTTACACACGCGTGTGCTTACGTTTGTGTTTCTCAACGTAAATTCATGCTACAGCGT	805		
679	accatccctctacacagagcccgaggtccatcgtcttaccctcatctgacaatttgcaaa	738		





Oy 414 ctgtctcagcttgcccccagagctcgaaattccaggtcttcaacaatgcccacatlgagag 473  
 Db 664 GGCTGTCCGGCTGGTCCCGA-----TGACATCGTATCCCGGAGAGATGAACGC 714  
 Oy 474 ctatgtgacacagacagcagctgagagctggaagctgacgtgtgtcagctagctg 533  
 Db 715 AACCAACGGCAGGAGCCCGAGAGCGGCTGAGTGAAGTGAATCCCATGTTGTGACCCCT 774  
 Oy 534 cctacacgccatctacagatgaggtctggtcttcatgcatgagttggtctgtgcccacta 593  
 Db 775 GCTCTCAGGAATCATTCAGTTTGGCTTAGTGTGTGCGACAGGTTTGGATTGTGGCCATATA 834  
 Oy 594 cctctccaggtcttcatctcaggggtcttcatgacggccgagctgacagatcttattc 653  
 Db 835 CCTCAGGAGCGCGCTGGTGGAGGTTTACACACCGCCGCCGCGACGCTTTCACATC 894  
 Oy 654 ggtgtcagaatcatctctgagctgacacatccctctcaacagggccaggttccatcgt 713  
 Db 895 CATGTTGAAATACCTGTTGGGTTAAAGCAAAAGCGTACAGTGGGATCTTTGCGTGTGT 954  
 Oy 714 cttaaccttcatgacattgcaaaacctcccccacacacatcgctgctgcatctt 773  
 Db 955 ATATAGTACACTGCTGTGTGCGAATGTTAAAAACCTCAACGTGTTCCTTAGCGCT 1014  
 Oy 774 cgtctcatcagcgtgtgctctctgtgtgtgtgagagagctcaatgctcgtacatgca 833  
 Db 1015 CGGCTGATGTTTGGTTTGGTGTGGTGGAGAGAGTTTATAGAGATTTTAAAGA 1074  
 Oy 834 caaatctgcttccccatccctcaacagagatgattgtgtgtgtgtgtgtgtgtgtgtgt 893  
 Db 1075 GAAATTTGCCAGACACCTCTCTAGAGTTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1134  
 Oy 894 cggggagctgtaagatgtgcccacaaagtatcacatcagatcgctgagagagacacagcgg 953  
 Db 1135 CGCGGGGTTTAACTTGCACGAGTCTTACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1194  
 Oy 954 gtcccaaccccggt 1013  
 Db 1195 GCTACTCCCTCCGCAACCGGACACGACGCTCTTCCACCTCGTGTGTGTGTGTGTGTGTGT 1254  
 Oy 1014 ctccctgacatctgt 1073  
 Db 1255 TGGCATGACGATCTGT 1314  
 Oy 1074 gcaagcagctgagctgagatctgaaacacagagatgagctgctgagctgacacatctt 1133  
 Db 1315 GCATGGCTACAGGTTGATGCAATCAGAGACTCATCGCTTGGGAGATATGCAACTCCAT 1374  
 Oy 1134 tgcctctctttaaataatctgcaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1193  
 Db 1375 CGGATCTCTTCCAGACCTTCTCATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1434  
 Oy 1194 tggagctgagagaaatcccaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1253  
 Db 1435 GGGAGACTGGAGGAAACACAGCTCGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1494  
 Oy 1254 catctgtctgt 1313  
 Db 1495 CATTTTACCCATGATTTCTTGTGAGTCAATTGCCCCAGGCTGTGCTCTCGGCCATTTGT 1554  
 Oy 1314 cgtgtcaatctcaagaatccctcaagaacacacagacacacacacacacacacacacac 1373  
 Db 1555 GATGCTGAACCTGAAGAGGATTTATGACAGTGTCTGACATGTGCCCCCTTCTGTGAGAAC 1614  
 Oy 1374 gagaagctgagacgt 1433  
 Db 1615 CAGAAATATAGCTGACCACTGTGCTTACCACTTTGTGCTTCCCTGCTGCTGCTGCTGCT 1674  
 Oy 1434 gacctatgt 1493  
 Db 1675 GGACTAGGAGT 1734

Oy 1494 gttcgaatgtgtatgtcactgtgcccaggtcatgtgacactgtacattgtgtaatccaa 1553  
 Db 1735 GAGTCCGAGTACCAAGGCTCTGGGCGACGCTCCGACACCATGATGATGATGATGATGATGAT 1794  
 Oy 1554 gacctataatagggcccgagatctccagggatgataaataatcactgactgtctccctc 1613  
 Db 1795 CGCATATGAGAGGTTGAAAGAAATTCCTGGAATTAATAATATTCAGATTAACGCCCCCAAT 1854  
 Oy 1614 ctacttgcacactcagagactctcaagcaaaaggtcatcgcccaagactg 1663  
 Db 1855 TTACTATGCAACAGTACTGTATAGCAAGCCCTAAAGAAAGACTG 1904  
 RESULT 6  
 LOCUS AY032863 2612 bp mRNA ROD 03-AUG-2001  
 DEFINITION Mus musculus chloride-formate exchanger mRNA, complete cds.  
 ACCESSION AY032863  
 VERSION AY032863.1 GI:15080863  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2612)  
 AUTHORS Knauf, F., Yang, C.L., Thomson, R.B., Mentone, S.A., Giebisch, G. and Aronson, P.S.  
 TITLE Identification of a chloride-formate exchanger expressed on the brush border membrane of renal proximal tubule cells  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (16), 9425-9430 (2001)  
 PUBMED 11459928  
 REFERENCE 2 (bases 1 to 2612)  
 AUTHORS Yang, C.-L. and Aronson, P.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-APR-2001) Internal Medicine, Yale School of Medicine, 333 Cedar Street, P.O. Box 208029, New Haven, CT 06520-8029, USA  
 FEATURES  
 source Location/Qualifiers  
 1..2612  
 /organism="Mus musculus"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 145..2352  
 /note="CFEX; transporter; possible ortholog of Homo sapiens SLC26A6"  
 /codon\_start=1  
 /product="chloride-formate exchanger"  
 /protein\_id="AAK51131.1"  
 /db\_xref="GI:15080864"  
 /translation="MGLORDYHVERPLNQEOLDLGHMGPAKTHQMTWFCNSRA RAHSLLOHVPVLMRLPRYPXREWLIDLLSGLSVAIMOLPGIATVALLGLPMEGL YSSEYPIVEFLFETSRHSISYGFPAVSVMSVSTESLTDKRAVOGLNATDARVO VAVTLISLNGFQVGLVHGFVPTLSLSEPLVSTYTAASVOVLASOLATVGVGKLS SHSGPLSVITVLEVCAQDLPEHGVTVYTAIVAGVALVTLAKNEKLRRLPIRPE LTLITAGTISTYGVKLNDRKRVNAGNITGLIPVAPKTELFATLVGNMFAIAVGF AIAVLSIKIFALRHRGVDSNOELVALGSLNIGGFQCPFVSCMSRSLSVOESTGN TOVAAGVASSFLILIIYKLGELFEDLPRKAVLAAYIYNLGMKMGFSDISLMKANV DLIILWTFVATILINDIGLAVISVLSLLVYRMLPHYSVIGVDPNDIYDVAE YSGAKVEPVGKVPERSATLVFANAEISDSLKKEGVDVRLITOKKRIJKOEMKIK RMKAKSKODASSTKTSVSVYNNYMLDPEVKSNDVGSRAKYNQCELDIYVSNNE DAKPFTISLSLGLPOGFRSLILDLSTLSVPTVCISIKNIIFRPRELEVETIA ACYSPVAOELAGHFDEFSITKQVFSVSDVAVFALSHKRSVSPVLAATKL"  
 BASE COUNT 542 a 725 c 701 g 642 t 2 others  
 ORIGIN  
 Query Match 15.7%; Score 355.8; DB 10; Length 2612;  
 Best Local Similarity 52.8%; Pred. No. 4e-65;  
 Matches 808; Conservative 1; Mismatches 693; Indels 27; Gaps 1;  
 Oy 112 cgcacatgctcagatgtctcagcccaagatcaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 171  
 Db 250 CGAACTGTTCGATGCTCCCGTCCGCGCCCACTTCTTCTGTGCTCAGACAGCTTCA 309

OY	172	gtgcctctcggcccccacaaagataaagaattaaagctatatactctccactgcgtgcgt	231
Db	310	gttcctgggctggttaccctccggatccttckccctgaaatggctctctgggctatctgttatct	369
OY	232	ggaactcagcggggatcccatcaggtccccaagagcatggaatttgcgtctgtgcacac	291
Db	370	ggccttgaactgaggcattatcagacacttccacagggcttg6ccattatccctttggcggga	429
OY	292	cttcctgcagtcgaatgagcctctactcctctcttcccccctccagcctaattctcgt	351
Db	430	ttggcctcctaatggtttggctgtacagatcttcttttaccctccgttctcatcttacttctgttt	489
OY	352	gggggggtgtccacacgtagcgcaaggttctgcggttatagaacatccctgggtgtgac	411
Db	480	ggatactccagacacatctctgttg6ggagaccttttgctgtatgtctgtatgttg6gacgt	549
OY	412	atctgtctgcagctgagcccgcaagatcgaattccaaagtcttcaacaattgcacaaatgag	471
Db	550	gtgacagaaatctctgacagcagataaagccttctgtcamaagggcttgaaatgcacacgt---	606
OY	472	agctatgtgaaacaaagcgaatggaagcttgagaagcttcaagtgtcagcttgcagctagcc	531
Db	607	-----GAAGATGACAGGGGTGACAGGTGACAGTGGCCATCACACTGACG	642
OY	532	tgacctcaccgcacatcatcagaatggagctgtgagcttcatgagtttgagcttbtgacac	591
Db	643	tttccctggctggcccttcttccagagtg6ggcttg6gcttg6tACACTTGCGCTTTCGTTGTCACC	702
OY	592	tacactctccagatctccatctccgggcttcgatbagcgccgcgagcttcgaatccctgat	651
Db	703	taccctgtacagacctgtggttgcggagctataccacggcgatcgaatgtgcaatgctctgctc	762
OY	652	tcgagtgtcgaagatcatctctcgagctgaccatccctctccacaaagcccaaggtctcatc	711
Db	763	tctcagcctcaagatgtgtgtttggatctaaactcagacgactcttg6ggcactctgtccctt	822
OY	712	gtctttaacctcatgtgacatttgcaaaaacactccccacacaaacatgcgcctgcgtcatc	771
Db	823	atctatctacagctgctgagagcttctg6ggcagctctccctgaacactgtg6cccg6acccgtgctc	882
OY	772	ctgcgtctcatcagcggctgctctctcgtgtgtgtgtgaagagctcaatgtctgcgtacatg	831
Db	883	acggcgattgtggcagagagtg6gcttggatctggtgaactgaatgaaagactgcgat	942
OY	832	cacaaagatctgccttccccatccctacagagatgtatgtgtgtgtgtgtgtgtgtgtgtgtgt	891
Db	943	cgccgctgtgccccctgccccatctcctggggaaatctctacacctatttg6ggcactctgataatt	1002
OY	892	tcggggggcgtgtaagatgccccaaagaatcatcatgtcagatgtgtgtgtgtgtgtgtgtgtgt	951
Db	1003	ttcttatgtgtgaaagctgaattgacacagatttcagagtgcatgtgtgtgtgtgtgtgtgtgtgt	1052
OY	952	gggtgtcccaaccccgcgtgtgcgt	1011
Db	1063	ggcgctaatccccggggtggcagcccaagacagacgtgttttgcaaacgttgtgtggaaatgccc	1122
OY	1012	ttcttcctagccatcggtgagctacgtcatcaacatgtgctatgtggcgcgaacctgtgcacac	1071
Db	1123	ttttccattgt	1182
OY	1072	aagcaaggtctagacgctgagttctgaaacccagagatgatctgcctctgcgtctgcagcaatc	1131
Db	1183	aggcatggctaacctgttg6gacagatcaaacagacacgttg7ag6cctttgg6cctcag7aaacctc	1242
OY	1132	cttgagccctctctttaaattcatatgcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1191
Db	1243	atttg6agagcttcttccagtgcttccccctgtgagctgtcctcatgttcttcg6agcttg6ttracg	1302
OY	1192	gatgtgagctgtaggagaaatcccaatgctgtgcacagcctgtgtgtgtgtgtgtgtgtgtgtgtgt	1251
Db	1303	gagagcagcgg6ag6gacacacacagaggttgctg6agcgtgatacttccctttttatctcttctt	1362
OY	1352	accaatgctgtgtcctgtagatcatctgtatcctctccctaaagtctgtgtcgtagaagccctg	1311

Db	1363	ATTATCGTCAAACTTGGGGAACCTTTCCGAGACCTGCCCCAAGGCCCTCTGCTGCTGTC	1422
OY	1312	atcgcgtcaaatctcaagaactccctcaagaactcaccgaccctactaccctgtgagag	1371
Db	1423	ATTATTGTGAACCTTAAGAGGCATGTATGAAGCAGATTCTCAGACATGCTGCTCTTTGGAA	1482
OY	1372	aagagaacgctggaactgtgtgatcctctgggtagtgtgagcttccctctctctctctcga	1431
Db	1483	GCMAACAGAGTGTGACCTGCTATCTGGGCTGGGAGACCTTGTGTGCCACATCTGCTGAA	1542
OY	1432	ctgcacctagtgtgtgagagtggtgcacctctcgcctcctcgtcgtgctgctcgaact	1491
Db	1543	CTGGACATTTGGCCTTGCACTTTCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1602
OY	1492	cagttcgaatgagctatgcatgcatggtcccaagtcagtcagtcagtcagtcagtcagtc	1551
Db	1603	CAGTGCCTCCATTTACTCCGCTCTCTGGGCGAGTGCCAGATAGGATTTATAGACGCG	1662
OY	1552	aagacctataatagggcccaagatattccaggggataataatcatacagctacgtccct	1611
Db	1663	GCAGATACTCTGGGGCCCAAGAGGCTCCGGGTGTGAAGTCTTCTGCTCTCTCTCTCT	1722
OY	1612	ctctactctcccaactcagaagctcag	1640
Db	1723	CTGTACTTCCGCAATGCTGAGCTCTACAG	1751

FLPVLPLSVGLSEMDLAASEIAVAVAIAVSVKVAATKDYITIDNGOEIAGISN  
VESEFSCPVATTAIRAEVSTGKTVQAGLSAVIYMAVIAKLEPLEPLOSVAL  
AAYIAINLKGMEQVOCVPRLMKRONKTDVAIVTPTCMSITILGDLGLAGLEGLIT  
VTVLVQVPSMNGLGSVPSTDIYKSTIKTKHNEEPEGVAKILEFSSPTFGANDGRKCA  
KSTYGFDAIRVYNKRLKALRIQKLKGGOLRAATKNGISDVGSSNNAFEDEVEEER  
EELDIIPKEILEIQVDMNSLEPVKVNVPVRPHSLTIVLLOCGAFADVGVRSRLAIYKE  
FORIDVNVYFALLIDVYLEKMEQCGFEDNIRKRFELTYVADAILTYLQNGAKSREGOD  
SLLEPTILIDOCCKDPLFMEAEINIEEELDVQDEAMRLRAS"

[illegible]

Db	1020	TGGGTTCTTGCCCTCCTGTACATGCGCGTGTGGGCGCTGTTTCTGACATGCTGGTGCCTTC	1079
Oy	1011	cttccctccatgccaatcgtagactaagatcaacaactcgtgctatagggccggagacccttggccaa	1070
Db	1080	CTTTTCATTTGGCCGTGGGGCTTTAGCGTAATTCAGATGTCGTAGGGAAAGTGTACGCCAC	1139
Oy	1071	caagaacagctaacgcgtgagatccgaacagagaatgactcgtctccgctcgtcagaaact	1130
Db	1140	CAAGCATGACTACATCATCATGATGGGACACGGAATTCATTCGCTTCGGGATTAAGCAAGCT	1199
Oy	1131	cttggctccctctttaaataatcaatgtaacttgcgtctgctccttcgtcactcgtgct	1190
Db	1200	CTTCTCTGGATTTTTTCTCTGTTTGTGGCTACCACTGCTGTCTGTCCGAACGGCTGTCCA	1259
Oy	1191	gagtagagctgagagaataatcccaagtgaggcaagccctgctgtgtcgtctcgtgtgtagatg	1250
Db	1260	GGAGAGCACTGGGGGGAAAGACACAGAGTGGCAGGCCCATCTCAAGCTGTATGTATGATG	1319
Oy	1251	caaccaatgctgctccctgaggagatcatctgatatcctcctccctaagctctgtagagaccct	1310
Db	1320	CGCGATTCGTGCCCTGGGGAAAGCTCTGGAAACCTTGAGAAAGTGTCTGTGGACACCGCT	1379
Oy	1311	gactcgtctcaatctccaagaactccctcaagcaactcaaccgaccctactactcgttggag	1370
Db	1380	CCTCATTTGGCAACCTGAAAGGATGTTTATGGAAGGTGTGATGTTTCTGTGTGGAA	1439
Oy	1371	gaagagcaagctcggagctgtgtcatctggttagtagactcctcctcctcctcctcctcag	1430
Db	1440	GCAGATATAGACTGATGCCGTTATCTGGGCTGTTTACATCATTAATGTCATCATCTTGTGG	1499
Oy	1431	cctgcacctatgctgtgacagtggtgtgcctctccgtcgtcgtgtgtgtgtcctccagac	1490
Db	1500	CGTGAGACCTCGGTTTGTGCTGGGCTGCTTTTATTTTGGACTACATGCTGTGCTGTGAGCT	1559
Oy	1491	tcaagttcgaatagtctatcagactcggcccaagttcatagagcaactgaatgaatgaatcc	1550
Db	1560	TGAGTTCCCTTCTGTCGATGCGCTTGGAAGTGTCCCCACACAGACATCTACAAAAGTAT	1619
Oy	1551	caagacctaataatagggccagagataccaagggatlaaataatcaatcaagtaactgtccccc	1610
Db	1620	CACACATTTTAAAAACCTTGAAAGAGCCTGAAGGGGTGAAGATTCGTGATTTTCCAGTCC	1679
Oy	1611	tctctacttgccaactcagagatctctcgggcaaaaggtca	1651
Db	1680	CATTTTATVGGCAATGTCATGTGTTTAAAAAATCGCTCA	1720
RESULT	8		
LOCUS	AF167411	3086 bp	mRNA
DEFINITION	Mus musculus pendrin (Pds)	complete cds.	
ACCESSION	AF167411		
VERSION	AF167411.1	GI:5802226	
KEYWORDS			
ORGANISM			
SOURCE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			



ACCESSION AY024359  
 VERSION AY024359.2 GI:14787222  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2235)  
 AUTHORS Adler, H.J., Merritt, R.C. Jr., Belyantseva, I.A. and Kachar, B.  
 TITLE Mus musculus organ of Corti mRNA for prestin  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2235)  
 AUTHORS Adler, H.J., Merritt, R.C. Jr., Belyantseva, I.A. and Kachar, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JAN-2001) NIDCD, NIH, Bldg. 36/5D15, 36 Convent Drive, Bethesda, MD 20892-4163, USA  
 COMMENT On Jul 17, 2001 this sequence version replaced gi:12658642.  
 FEATURES  
 Location/Qualifiers  
 1..2235  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /tissue\_type="organ of Corti"  
 1..2235  
 /codon\_start=1  
 /product="Prestin"  
 /protein\_id="AAG59999.2"  
 /db\_xref="GI:14787223"  
 /translation="MDHAENEIPAEIORYVERPIFSPVLOERLHVKKDVTESIG  
 KLRQFTCPKIKINIIYMLPIPKMLPAKPKKRYLDLVSGISTGYLOLPQCLAF  
 MIAVPPYEGISSTPYIMKCFPTSHSISIGPAVSLMIGAVNLPDDIYIG  
 GVNAITNGEADALRVKAVSVTLISGLIIOCLGCRGPAIATLEPLVGFPTAA  
 VHVETSMILKLVGKTRKSGISFVSVYTAVALNVKNLSLGVLMVGLLGK  
 EFNRFEXKLPAPIDLEFAVAGTSGISAGNHLSEYVDVGTPLILPANDTS  
 LEHLYVDALIAIVGFSVTISMAKTLANKGYODGNOELIAGICNSISLPOTS  
 ISCSISLSVOEGTGKQIAGCLASMLIIVITATGPLESPISPAVISAIVYLNK  
 MEMQSPDPFTWRTSKIEPTIWTTFPSLSLELDYGLITAVITALLVITRTSPSI  
 KVLQGLPDTVIDIDATEVKEIPGKIFQINAEVRYANSDDISSALKRRTVPAL  
 IMGARRRAMRYAKEVGNANVAVVDAEVDGEMTKDEEDDEVKFPPIVITKT  
 FPELQRLPQGENVHTVILDTQVNFVDSGVGTLAGIVKEGVGIGIYVLAGSPQ  
 VVNDLTRNFPENPAKELFPHSHIDAVLGSQVREMAEQEATSLPDEMDENATPT  
 TPEA"  
 BASE COUNT 587 a 516 c 542 g 590 t  
 ORIGIN  
 Query Match 14.9%; Score 337.2; DB 10; Length 2235;  
 Best Local Similarity 51.7%; Pred. No. 3; Se-61;  
 Matches 821; Conservative 0; Mismatches 758; Indels 9; Gaps 2;

436 tcgaattccaggtcttcaacaatgcacccaatgagagatctgtgacacgaagcagatg 495  
 463 GATA-----TTGTCATCCAGAGAGAGTAATGCAACCAACGCGG---ACAGAGCCAGA 513  
 496 gaggtcagagagctcagctgtcagctacgctagcctcagctcagccatccagatg 555  
 514 GATGACATAAGAGTGAAMATCCCATGTCGTACCTTACTTACAGGAATCAATGATTT 573  
 556 ggtcgtcagctcagctcagctgtcagctgtcagctcagctcagctcagctcagctc 615  
 574 TCCCTAGGTGTGTAGCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 633  
 616 ggtcctacagagccgagccgagctcagctcagctcagctcagctcagctcagctc 675  
 634 GCGTTTACACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693  
 676 ctgaccatccctcctcactacagagccagaggtccatcgtcttcaactcattacattgc 735  
 694 GTCAAAACAAACCGGTACAGTGAATCTTTTCAGTGTGTATAGTACACTTCTGTGTG 753  
 736 aaaaaccccccac 795  
 754 CAGAAATGTTAAACCTCAACCTGATGCTTCCCTAGGCGTGGCTGATGTTTGGTTTG 813  
 796 ctggtgtcgtggaagagctcagctcagctcagctcagctcagctcagctcagctc 855  
 814 CTTGTTGGGTGAGAGAAATTAATAGAGATTTAAAGAGAAATTTGACAGCACCATTCT 873  
 856 aagagatgattgt 915  
 874 CTAGATTTCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 933  
 916 aagttcactcagatcgt 975  
 934 TCTTACAGT 993  
 976 gtgtgtcagatggaagacatgaaagacagctcctcctcagcagctcagctcagctc 1035  
 994 GACACCAGCCTTCCACCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1053  
 1036 gtcacacacgt 1095  
 1054 TCAGTGACGATTCATGCGCAAAACCTTGCAAAATTAACATGCGATTCATGCGCG 1113  
 1096 aaccagagatgatcgt 1155  
 1114 AATCAGAGACTATTGCTTGGGATATGCAACTCCATTGGATCTCTTCCAAACCTTC 1173  
 1156 gtcatttctgt 1215  
 1174 TCGATTTCCTGCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1233  
 1216 gtgtgtcagctgt 1275  
 1234 CTTCGAGGT 1293  
 1276 ctgtatccctccctcactcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1335  
 1294 TTTGAGTGTATACCCAGGCTGTCTTTCGCCCATTTGTATGTGTACCTGAAAGAAATG 1353  
 1336 ctcaagcaactacacgacccctactactcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1395  
 1354 TTTATGCAATGCTCAAGACTGCTGCTTTTCTGTGAGAACAGCAAAATAGAGCGACATC 1413  
 1396 tgggt 1455  
 1414 TGGCTGACCACTTGT 1473  
 1456 gtgcgtcctcgtcgtcgt 1515  
 1474 GTGATCATTTGCTCTGCTCAAGTGTATTTATAGAACACAGAGTCAAGCTCAAAAGTCT 1533



Db	Accession	Source	Location/Qualifiers
Db	1141	CTTGGAACCTTAAGGGAATGCTATGCAAGTTTGGTGGAGATAGGACAGACTGTGGAGAAG	1200
Qy	1375	agcaagctgagctctgtgcatctgaggtgagtgagctctcctcctcctcctcctcctg	1434
Db	1201	GACAAATACGACTGTGTTAATTGATGATGACACTTCACTTTGGCATGTGCTCGGACTA	1260
Qy	1435	ccctatggtgtgagcagtggtgtgtcgcctctcgcctcgtgctgagtccttcacagctag	1494
Db	1261	GGATTTGGCTCTGGCGGCAAGGTGTGGCGTTTAAACCTCCACCATCGTATTCAGGACTCAG	1320
Qy	1495	ttctgaaatgctatgagcagctgcccaggtcatgagcagctgacattatgtaatcccaag	1554
Db	1321	TTTCCAAATGACACCACTGTGCTAATCATCGAGAGGACACACATCTATAAGATGAGAAA	1380
Qy	1555	accta 1559	
Db	1381	GACTA 1385	
RESULT	11		
LOCUS	RNO303372		
DEFINITION	Rattus norvegicus mRNA for prestin.	ROD	29-MAR-2001
ACCESSION	AJ303372		
VERSION	AJ303372.1	GI:12188917	
KEYWORDS	prestin.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 2235)		
AUTHORS	Ludwig, J., Oliver, D., Frank, G., Klocker, N., Gummer, A.W. and Fakler, B.		
TITLE	Reciprocal electromechanical properties of rat prestin: The motor molecule from rat outer hair cells		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4178-4183 (2001)		
MEDLINE	21173705		
REFERENCE	2 (bases 1 to 2235)		
AUTHORS	Ludwig, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JAN-2001) Ludwig J., Physiologie II / MNF, Universitaet Tuebingen, Ob dem Himmelreich 7, Tuebingen D-72074, Germany		
FEATURES			
source	location/Qualifiers		
	1..2235		
	/organism="Rattus norvegicus"		
	/db_xref="taxon:10116"		
	1..2235		
	/codon_start=1		
	/product="prestin"		
	/protein_id="CAC21555.1"		
	/db_xref="GI:12188918"		
	/translation="MDHAEEKEIPAEFTOKYVERIPFSPVLOERLHVKNKVDYSIGF		
	KLAAPFTCPKRVNIIYMLPIKMLPAYFKEVIGADISGVLOLPQGLAEFPEFSS		
	MLAAPVYGVSSFPYIMYICFPTGSHIFISGFAVSLMGVAALVDDYIVIGG		
	GVNNTGTEARDALRVKAVMSVTLISGLIOCLGRCGFVAITLPELVGFTAAAG		
	VHVTSMKLYFGVTKTKYSGIFSVYATVGVNKNKNCISGVGLMVGGLLGGK		
	ENRERKEKLPAPIPIEFERFANVMGTGISAGNLEHSYVDVGTLPGLLPANPDPS		
	LFHLVYDAIAIVGFSVTISMATKANRGVYVDGQDELIALGICNSIGSLCVTSS		
	LSCASRLVDEGTGKQLAGCLASMLILVILVATGELFESLPQAVLSAIVYNLKGK		
	MFMOFSDLPEFMWTSKIELITLMTTPVSSLEIGDYGLITAVIALLTVIRTPSSY		
	TVLQDLPDPTDYIDIDAAEEVEKEIPGIFQINAPIYANSDLASALKRKTGNPAIP		
	IMGARRKMRKRYAEKAVGNNANNTAANVAVDAEVDGNAATKPEEDDEVKPPYIKTKT		
	EPFELQRLPDGGENIHYYIIDLTPQVNRFDSDSVGVTTLGAIYKEYGDVGIYVLAGCSAQ		
	VVNDLTSRFFENPALKEELRHSIHDAVGSQVRAMAEQETVLPQEDMEPNATPT		
	TPEN"		
BASE COUNT	560 a 538 c 563 g 574 t		
ORIGIN			
Query Match	14.7%; Score 332.8; DB 10; Length 2235;		
Best Local Similarity	51.5%; Pred.No. 3e-60;		

[illegible]









Oy	629	ccgcggccctgcgaatccctgatttcgggtctcaagtaatcttcggagcagaccatccct	688
Db	901	ctgcctcccttccaaagctctgctctgcacagctaaagatttctcctaattgtttcaaccamaa	960
Oy	689	ccctacaaagcccccagggtlcatcgtctcttcaacttcatagatcttgcataaaacccctccc	748
Db	961	actacaaatgagattctctctattattatctatccgtgggttagattttttcaaaatattggng	1020
Oy	749	aacaccacatcgccttcgctctatcttcgtctctatcaaggggtgccttccgtgctcgtgta	808
Db	1021	ataccaaattcttgcatatttctacgtctcgatttgctcacattgctgtctgtataggcagtta	1080
Oy	809	agagagctcaatctgcgtctacatgcaagaatcgtctcccatccctacagatcgtatg	868
Db	1081	aggaatttaattgatcgggttttagacacaaatcccaattcccttatctctatnagaattatgg	1140
Oy	869	tgtgtgtgtgtgcacaagctatctccgggggtcgttaagatgcccataaagatcatcatgc	928
Db	1141	tgacagataattgctnactgcatcttccattgatnaggaccacaactggamaaaaattttacattgctg	1200
Oy	929	agatcgtgtggagaaatctcaacgcgggtttcccccaccccggtgcgctgtgtgtctcaagct	988
Db	1201	gcatttgttaattccattcccaagggggttttggcctcttaacctttccaccctgtgagcttgt	1260
Oy	989	gaaagagacatgatagagcacagcctctccctaaacatcgtagacatcagatcaaacctgcg	1048
Db	1261	ttctcgagatagctgctgcatcatattttccattgcctgtgggtggcttattgctattggcaggt	1320
Oy	1049	ctatggtgcggagccctgcgccaagaacaggtctagagctgtgattctcgaaacagagatga	1108
Db	1321	cagtaggaaaaatgatattatgccaccaaagtatgatattaccatgcattgggaacacggaaattca	1380
Oy	1109	tcgctcccggtctgagaaactctcttgcctctctttaaattcaatcatgtctgtg	1168
Db	1381	ttgccttttgggatcagcaaatctttctcagagattcttcttctgtttttgttgccaccactgc	1440
Oy	1169	cgcttcttgtaactctgcgtctgtgatatgagctgtgagaaatccaggtgcagcagctgt	1228
Db	1441	cttttttccccacggcgctgcagaaagacactgaaagaaacacacaggttgctgcatnca	1500
Oy	1229	gtgtgtctctgcgtgtgatatcaacatgctgtgctctgtggatctatctgtatctctcc	1288
Db	1501	ttctctgctgcgattgtgatattgcccattcttggccctggggaagcttctggaaaccttgc	1560
Oy	1289	ctaaagctctgtctagagccctgtatcgctgttcaatctcaagaacctccctcaagcaactca	1348
Db	1561	agaaagtgggtcttgccagcagctgttgtaattggcaactgaaaggagttttatgcagctgt	1620
Oy	1349	ccgaagccctactaactctgtgtgagaagaagaagctgtgactgtgtgcaactctgtgtagtga	1408
Db	1621	gtgacatttccctgcctgtgtggagacaaataaagattgatctgtatnctggggtgtttacgt	1680
Oy	1409	tcctctctctctctctctctcactagcctgcctcatggtgtgtagtgcgtgtgtgcctctcc	1468
Db	1681	gtatagtgctccattcatcttctggggcttgatctgcggtttactactctggccttatattttggac	1740
Oy	1469	tcctgtgtcgtgctctccagaactcagtttcgaatggtcatgacgttggcccgagtgcatgcg	1528
Db	1741	tgtttgactgtgtgctctgaagattcacttttcttcttggaaatggcctttggaaacatcccta	1800
Oy	1529	acactgacatttatgtgaatcccaagaactataatagggccagagatalccagggatla	1588
Db	1801	gcacagatattctctcaaaaagacaaatattacaanaaacattgaagacctcaagagatga	1860
Oy	1589	aaatcatcaatgatactgctccctctctactcttgcacactcagaatcttcagggcaaaag	1648
Db	1861	agatttcttaatttttccagctcattttttctatggcaatgtcgatgttttataaaaattgta	1920
Oy	1649	tcatacgcca	1657
Db	1921	tcgaagtcca	1929

RESULT	14
LOCUS	HUMDRA 2881 bp mRNA PRI 31-DEC-1994
DEFINITION	Homo sapiens colon mucosa-associated (DRA) mRNA, complete cds.
ACCESSION	L02785
VERSION	L02785.1 GI:291963
KEYWORDS	colon mucosa-associated protein.
SOURCE	Homo sapiens colon cDNA to mRNA.
ORGANISM	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. Schweinfest,C.W., Henderson,K.W., Sister,S., Kondoh,N. and Papas,T.S. 1 (bases 1 to 2881)
TITLE	Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (9), 4166-4170 (1993)
MEDLINE	93348250
FEATURES	Location/Qualifiers
source	1..2881
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_type="epithelial cell"
	/tissue_type="colon"
	185..2479
gene	/gene="DRA"
	185..2479
CDS	/note="Nuclear localization signal at AA 569-573, 576-580, 579-583; acidic transcr. activ. domain 620-640.; homeobox motif 653-676" /codon_start=1 /protein_id="AA58443.1" /db_xref="GI:291964"
	/translation="MEPEFGNQYIVARPVSTNAFEBNRKTKGRHKFTPLDLKCCS /CDPKARIVLSLEPIASWLPAYRLKEMLSDISGISTGIVAVLOGLAFLALVDIP VYGJIAFFPAIIYLFEPTGSRHISVGPFILSMWGVAAGVASKVPDRNATTLGLP NNSNSSLLDDERVRVAASAASVYSLIDLAGLIRIGVVLYLESLSIGTTAA VHVYSOLKIFPOLIYVSHDPVSIFNVLSVPSQIEKNADIALVALIVLVSVIK EINORFKDKLPVPIPIEFITVIYAAGSYSCDPRKNRKAADVDMNGPPIETPEVE TFQWTGDCFGIAMVAFVAVSAVSISSKTDPLUGNDLPLAKGSLNYCVGFRCGA GSAFASRSAAVOESTSGCKTQIAGLIGAILVLIVLAGFLAPLKSVLALAAGNLKG MLMOFAELGRLMRKRDYDLIMTMTEFIETLIVLAGLASAVFOLLTIYFRPOFC STLANRTNIYKNKKDYDEYPEGVKIERPCSPFYFANIGFERKRLDAVFSPLR ILRRNKRLKRIRLOKGLGVTPKGICFTVTPTIOSDELDNNOIEVLDOPIINTD LPHRIDMDPLPIETPKSLSHLSDPSAASFIDVSSYRGKSLIOERIRKNDY LPTHDDEIKLNREFFEDGVKSIFETLIHDVALIHLMKDYSTSKRNPSOEKGG IDFTININGLNRKVEVEPETKF"
BASE COUNT	839 a 578 c 596 g 868 t
ORIGIN	
Query Match	14.4%; Score 326.2; DB 9; Length 2881;
Best Local Similarity	51.7%; Pred. No. 7.2e-59;
Matches 798; Conservative 0; Mismatches 733; Indels 12; Gaps 2;	
OY	124 agatgttcctcaagcagaatcaaaagtctgtagtttgaggctgcgcctcgctgctctctgg 183
Dd	314 AGCTGTTCGCCAACAAAAGGCCCAAGAGAATGTCTCTCTTTGTGTCCCATAGCATTCTGG 373
OY	184 ctccccaagtaacaagattaagaactacatcatctctgacctgcgcgtgagactcaagcgg 243
Dd	374 TTGGCAGCATATCGCGCTTAAMAGANTGTGTGCTCAGTGATATTTGTTCGTGGTATCAGACA 433
OY	244 ggatcacatcgaagtcaccaagaagcattgcatctgctctgcgcgaacctctcgaagtc 303
Dd	434 GGGATTGTGGCCGTACTACAAAGCTTTTGACATTGECTCTGCCTGCTGACATTCGCCCGCAGTC 493
OY	304 aaibgcctctaactctctctcccccctctgaccactctctccggggggggtgtcac 363
Dd	494 TATGGGTGTATGCATCCCTTTTTCCCAGCCATAATATCACTTTTCTTGGCACTTCCACA 553







```

Db 144 --ALNDSMINETARDAAARVOASTLSVLVGLFQVGLGILHGFVVTYIASEPLVRGYTTAA 201
Qy 212 GLQILISLVKTYIFGLTIPSYTPGSIYFTFIDICKNLPHNTIASLIFALLISGAFLVYKE 271
Db 202 AVQFVFSQKTVFGLHSHSGPLSLITVLEVCCKLPQSKVGVVTAAGVAVLVVYKL 261
Qy 272 LNAIRMKIRPPIPTTEMIVVAVATASISGCKMPKKYHMOIVGEIQRGFPPTVSPVVSQWK 331
Db 262 LNDLQOOLPMPPIGELLTLTGATGISYGMGLKHFEDVDVGNIPAGLVPPVAPNTQLFS 321
Qy 332 DMIGTASLAVSYVINAMGRITLANKHGVDSNOEMIALGCSNFGSPFKIHVICAL 391
Db 322 KLVSAFTIAYVGAFAISLIGKIFALRHGRVDSNOELVALGSLIGITQCFPVSCSM 381
Qy 392 SVTLAVDAGAGKSOVASLCSVLVYMTLVGLIYLPPLKSVGLALLAVNLKNSLKQLTD 451
Db 382 SRSIVQESTGNSQVAGAISSFLITLITVLEVCCKLPQSKVGVVTAAGVAVLVVYKL 441
Qy 452 PYLIMRKSILDCCTWVVSFLSFLSLPYGAVAGVAVSVLVVVFOTOPFRNGTALAQVMDT 511
Db 442 MRSIMKANRADLLIMLVFTATILNLNDLGLVAVIFSLLVVVTQMPHYSVLGQVPT 501
Qy 512 DIYVNPYTNRAODIOGKITITCSPLYFANSEIF-----ROKVIK 553
Db 502 DIYDVAEYSAKEVKGKVRSSATYVFANAEFYSDALKRCQGVDFLSQKKKLK 561
Qy 554 --TVSLOELQOD---FENAPPTDP-----NNNQTPANGTSVSYTFSPDS 593
Db 562 QEOLTKLOLQKEEKLKROAASPKGASVINYNTSLEDKRSNNVEDCKMKQVS---SGDK 617
Qy 594 SSPAOSPEPAEAEPGESDMLASVPPVTFHTLILDMSGSVFVDMKIRKALKLSITYG 653
Db 618 MEDATANGQEDSKAP-DGSTLKAIGLPPDPFHSILDLGALSFDVTCCLKSLKNIHFDR 676
Qy 654 KIGKVFVLVNIHAQVYNDISHGVFEDGSLCKHVPFSDIHDAVFAQANMADV 706
Db 677 EIEVEVYMAACHSPVVSQLEAGHFF-DASITKHLFASVHDAVTFALQHPRPV 728

```

```

RESULT 2
Q9BXK9 PRELIMINARY; PRT; 759 AA.
ID Q9BXK9;
AC Q9BXK9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE SOLUTE CARRIER FAMILY 26 MEMBER 6.
GN SLC26A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21145584; PubMed=11247665;
RA Waldeger S., Moschen I., Ramirez A., Smith R.J., Ayadi H., Lang F.,
RA Kubisch C.;
RT Cloning and Characterization of SLC26A6, a Novel Member of the Solute
RT Carrier 26 Gene Family.
RL Genomics 72:43-50(2001).
DR EMBL: AF288410; AAK19153.1;
SQ SEQUENCE 759 AA; 82966 MW; 63CB0B756C9675C6 CRC64;

```

Query Match 32.5%; Score 1258; DB 4; Length 759;  
 Best Local Similarity 36.6%; Pred. No. 1e-70;  
 Matches 261; Conservative 162; Mismatches 238; Indels 52; Gaps 8;

```

Qy 32 PVGEKLNARFCSAKIKAVFGLPYLVSLPKYKIDYIIPDLGLSGSIGVPGMA 91
Db 51 PRHQMWTWQCSRAARAVALLQHLPLVLMRPRYPRVDMGLGLSLGSAVIMLPGSLA 110
Qy 92 FALLANLPAVNGLYSSFFPLITFFLGLGVHOMVPGTFAVISILVGNICLOLAPESKQV 151

```

```

Db 111 YALLAGLPVFGVGLSSFFVPEIYFLFTSRHISVGTFAVMWSMGVSTESLAPQ----- 164
Qy 152 NNATNESYVDVDAAMEELHVSATLACTAILOMLGDMOGFVAVIYSEFIRGEMTAA 211
Db 165 --ALNDSMINETARDAAARVOASTLSVLVGLFQVGLGILHGFVVTYIASEPLVRGYTTAA 222
Qy 212 GLQILISLVKTYIFGLTIPSYTPGSIYFTFIDICKNLPHNTIASLIFALLISGAFLVYKE 271
Db 223 AVQFVFSQKTVFGLHSHSGPLSLITVLEVCCKLPQSKVGVVTAAGVAVLVVYKL 282
Qy 272 LNAIRMKIRPPIPTTEMIVVAVATASISGCKMPKKYHMOIVGEIQRGFPPTVSPVVSQWK 331
Db 283 LNDLQOOLPMPPIGELLTLTGATGISYGMGLKHFEDVDVGNIPAGLVPPVAPNTQLFS 342
Qy 332 DMIGTASLAVSYVINAMGRITLANKHGVDSNOEMIALGCSNFGSPFKIHVICAL 391
Db 343 KLVSAFTIAYVGAFAISLIGKIFALRHGRVDSNOELVALGSLIGITQCFPVSCSM 402
Qy 392 SVTLAVDAGAGKSOVASLCSVLVYMTLVGLIYLPPLKSVGLALLAVNLKNSLKQLTD 451
Db 403 SRSIVQESTGNSQVAGAISSFLITLITVLEVCCKLPQSKVGVVTAAGVAVLVVYKL 462
Qy 452 PYLIMRKSILDCCTWVVSFLSFLSLPYGAVAGVAVSVLVVVFOTOPFRNGTALAQVMDT 511
Db 463 MRSIMKANRADLLIMLVFTATILNLNDLGLVAVIFSLLVVVTQMPHYSVLGQVPT 522
Qy 512 DIYVNPYTNRAODIOGKITITCSPLYFANSEIF-----ROKVIK 553
Db 523 DIYDVAEYSAKEVKGKVRSSATYVFANAEFYSDALKRCQGVDFLSQKKKLK 582
Qy 554 --TVSLOELQOD---FENAPPTDP-----NNNQTPANGTSVSYTFSPDS 593
Db 583 QEOLTKLOLQKEEKLKROAASPKGASVINYNTSLEDKRSNNVEDCKMKQVS---SGDK 638
Qy 594 SSPAOSPEPAEAEPGESDMLASVPPVTFHTLILDMSGSVFVDMKIRKALKLSITYG 653
Db 639 MEDATANGQEDSKAP-DGSTLKAIGLPPDPFHSILDLGALSFDVTCCLKSLKNIHFDR 697
Qy 654 KIGKVFVLVNIHAQVYNDISHGVFEDGSLCKHVPFSDIHDAVFAQANMADV 706
Db 698 EIEVEVYMAACHSPVVSQLEAGHFF-DASITKHLFASVHDAVTFALQHPRPV 749

```

```

RESULT 3
Q9EPH0 PRELIMINARY; PRT; 744 AA.
ID Q9EPH0;
AC Q9EPH0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE "The reciprocal electromechanical properties of rat prestin: the motor
DE PRESTIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Ludwig J., Oliver D., Frank G., Kloecker N., Gunmer A.W., Fakler B.;
RT "The reciprocal electromechanical properties of rat prestin: the motor
RT molecule of rat outer hair cells."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ303372; CAC21555.1;
DR InterPro: IPR002645; STAS.
DR Pfam: PF00916; STAS.1;
DR Pfam: PF00916; Sulfate_transp.1;
SQ SEQUENCE 744 AA; 81278 MW; E49E842CF7A3CD58 CRC64;

```

Query Match 32.5%; Score 1255.5; DB 11; Length 744;  
 Best Local Similarity 35.4%; Pred. No. 1.4e-70;  
 Matches 265; Conservative 163; Mismatches 267; Indels 53; Gaps 11;

QY	7	RYVVDRAAASLTLEFDEDEKEDR-TYVVGKRLNAFCSAKIKAVFGLPLVLSMLPKY	65
Db	15	KYLVERIPFSHPVQOEKLHKDKKDTSDIGKLRQAQFCTKKRKNITMLPLTKMLPAY	74
QY	66	KIKDYIIPDLGLSGSIOVPOGMAFALLANLPAVNGLSSEFPLTYFFLGVHOMVP	1255
QY	126	GTFVAVSILVGNICLOLAPES-KQVYNNATNESYVDTAAEAEHLVSAITLACTRAITQ	184
Db	135	GPFVAVISLMTIGVAVRLVPDPDIVIGVMAITN-----GTEARDALIRKVAASVTLSSGIIQ	1900
QY	185	MGLEFMFOGFAVIAIYSESFIRGFTAGLQILLSVLKTYIGLTFPSYTGSGSYFFPDI	2444
Db	191	FCLGVCRCRGPFAIITLTPYLRGFTTAANAHVFTSMKLTLEGVATKRSGLEFSVYISVAV	2500
QY	245	CKNLPHNTNIASL-----IFALISGAPLVLYKELNARYMHKIRREPIPTMIIVVAVTAISG	2399
Db	251	LQNKNNINVCISIGLAWFGLLGG-----KEFEREKELPAPIPEFAVAVMGTISA	3050
QY	300	GCKRPKRYHQIYVEIQRGFTTPVSPVYSQMKMDIGTAFSLATSYIYNALMRTLANKH	3599
Db	306	GFNLHESYSDVVVGTPLTGILPPRANPPTSLFHLVYDAIAIAIVGSFVTTISMAKTTLANKH	3655
QY	360	GVDVDSNQEMIALCSCNFSGFEKIHVICALSVTLAVDAGGKSQVASLCVSLVAVITM	4199
Db	366	GLOYDGNDELALALCISIGSLFOTFSTISCSLSMISVQEGTGKQTQAGCLASIMILVIY	4255
QY	420	LVLIQIYLYLPKRSYVAGALIAVNLKNSIKQLTDPYLYMRKSKLDCCIVWSFSLSEFSLP	4799
Db	426	LATFLEFESTLPQAVLSAIVNLKGMFMQESDLPFFMRTSKIELTILWTLTFVSSLEFLGLD	4855
QY	480	YGVAVGVAFSVLVVPQTOFNGFALAOVMDTIDIVMPKRYNNRQDIOGKIITYCSPLY	5399
Db	486	YGLTVAIVAILTYIYRQSPSYTVLQQLPRTDVIYIDIDAEVKEKLPGRKIPQINAPIY	5455
QY	540	FANSEIFROKVIANT-----VSLQELQODENAPPTDPNNQTPANQTSYVI	5875
Db	546	YANDSLVSSALKRKTGVNPAIIMGARRAMKRYAKEVNA-----NINATVVKYDAE	5985
QY	588	TFSPDSSSPAOSE-----PPASAEAPGEPSSMLASVPEPYTFHLLIDMSGVSFVDLMGI	6424
Db	599	VDGEAKRKPEEDEVEKFPPIVITTT-FPELEORFLPOGENIHRVILLDFQVNMDSVGV	6575
QY	643	KALKMLSTSYKIGVKEVFLVNIHQVYNDISHGCVFEDGSLCKCHVFPSTHDAVLFQAN	7024
Db	658	KTLAGIYKEIYDVGIIYTLACSSAOVNDLTSNRFENPALK-ELHRSHIDAVL-----	7111
QY	703	ARDYTPGHNFQAGFADAEALSLYDSEEDI	730
Db	712	-----GSQVREMAAEOETTVLPPQEDM	733
RESULT	4		
Q9JRK02			
AC	Q9JRK02	PRELIMINARY;	PRT: 744 AA.
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	PRESTIN.		
GN	PRESTIN.		
OS	Meriones unguiculatus (Mongolian jird).		
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Gerbillinae;		
OC	Meriones.		
OX	NCBI_taxid=10047;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=20279283; PubMed=10821263;		
RT	Zheng J., Shen W., He D.Z.Z., Long K.B., Madison L.D., Dallos P.,		
	"Prestin is the motor protein of cochlear outer hair cells."		

RL Nature 405:149-155(2000).  
 DR EMBL: AF230376; AAF1715.1; -.  
 DR InterPro: IPR002645; SRAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; SRAS\_1  
 DR Pfam: PF00916; Sulfate\_transp\_1.  
 SQ SEQUENCE 744 AA; 8118 MW; 1CDF6589BDACB8C CRC64

Query Match	32.1%	Score 1242.5	DB 11	Length 744
Best Local Similarity	35.4%	Pred. No. 9.2e-70		
Matches 263	Conservative 159	Mismatches 278	Indels 43	Gaps 10

QY	7	RYVVDRAAYSLTTFDEDEFEKKDR-TYVVGKRLINAFECSSAKIKAVFGLLPVLSMLPKY	65
Db	15	KYHERETIFSHPIVQDELHMKDKVSEISGKLQCAFCTPKTKIRNITMYLPTTKMLPA	74
QY	66	KINDYIIPDLGLSGSIOVPOGMAFALLANLPVANGLYSFFPLITYFFLCGVHOMP	125
Db	75	KFEKEYDLVSGISTGVLLPOGLAFAMLAAPVPEGLSSEFPVIMYCFEFTSRHIS	134
QY	126	GTEFVILVGNICLOLAPES-KOVFNMMNPNESYVDFAEMAEERHVSATLACIATIQ	184
Db	135	GPFVILSLMGVAVRLVPDVIYIGGVNMTN-----GTEARDALRKAASVLLSGTIQ	190
QY	185	MGLEFMQFGEFAIYLSSEFIRGEFTMTAAGLOILISVLKIFGLIPSYTGCISVTEFIDI	244
Db	191	FCGLGVCRCRFAIYLEPDLVRGFTTAANVHYTSMKLTYLGVTAKRRSGLEFSVYSYAV	250
QY	245	CKNLPHNTIASL-----IPALISGAPLVLYKELNARYMKHRIKRPPIEMIVVVAATISG	299
Db	251	LQNKNNLNVSISGLVWVFGLLILG-----KEFNERKEKLPAPIDLEFPAVWGTGISA	305
QY	300	GCKMKPKYHMOIYVEIQRGPTPVPSPVYSOMKIDGAFSLATISYTNLAMPGRITLANKH	359
Db	306	GFNLHESYSDVGVTEPLGLPLPAMPOTSLEHLVYVDAIAIAYGSEVYTTISMAKTLANKH	365
QY	360	GYVDVDSNQEMIALGCSNFFGSFFKIHVICALSVTLAVDAGGKSQVYASLCVSLVNIYM	419
Db	366	GYVDGNOELTALGICNSIGSLFQFTSISCSLSRSLVOEGGKTKOAGCLASIMILLVI	425
QY	420	LVLDGIYLPKPSVYLGALIANVNLKSLQKLDPIYLYMKRKSLOCIMVWSFSLSP	479
Db	426	LATPFLFESLPQAVLSAIYVNLKGMFMQSPDLPFEWRTSKIELTLLTTFVSSLEGLD	485
QY	480	YGVAVGAFSVLVVPQTOEFNGFALQVMDTDIYVMPKTYNRAODLOGIKITTYCSPLY	539
Db	486	YGLTAVIATIALITYYTRQTSXYVLDQLPRTDVIYDIDATYEEVKELIPGIKIOTIQAIPY	545
QY	540	FANSEIFROKVIAT-----VSIOLOQDEENAPPTDPNNNOPTPANGTSYSYI	587
Db	546	YANSDLYSNALKRKTGVNPAIMGARRKRAMKYAKEVGA--NIANAAYVAKVGEVDGEN	603
QY	588	TFSPDSSPAPOSEPPAPASAEARGESDMLASVPPVYTHLTLDMSGSPFDMGICIRLAK	647
Db	604	ATKPEEEDDEVKYKPIYIKTT-FPEELQREMPQOTENVHTTIIIDTQVNFIDSQVKTLLAV	662
QY	648	LSSTYGRIGVYFVLVNIHAOYVNIISHGVEQDSLECKHFPFSDHAPVFAQANARDYT	707
Db	663	MYKEYGDVGIYVYIAGCSPOVYVNDLTRNRFFENFALK-ELLFHSIHAYVL-----	711
QY	708	PGHNFQCAPGDALSLYDSEEDI 730	
Db	712	GSHVREAMARQEAASAPPQODM 733	
RESULT	5		
Q99NH7	099NH7	PRELIMINARY; PRT; 744 AA.	
DT	01-JUN-2001	(TREMblrel. 17, Created)	
DT	01-JUN-2001	(TREMblrel. 17, last sequence update)	
DT	01-JUN-2001	(TREMblrel. 17, last annotation update)	

DE PRESTIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_Taxid=10090.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=ORGAN OF CORTI;  
RA Adler H.J., Merritt R.C.Jr., Belyantseva I.A., Kachar B.;  
RT "Mus musculus organ of Corti mRNA for prestin."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY024359; AAG5999.1; -  
SQ SEQUENCE 744 AA; 81380 MM; 56E842748288E96F CRC64;

Query Match 32.1%; Score 1242; DB 11; Length 744;  
Best Local Similarity 36.5%; Pred. No. 9.9e-70;  
Matches 262; Conservative 154; Mismatches 260; Indels 42; Gaps 10;

QY 7 RYVVDRAVYSLTLEDFDEFEKDR-TYPYGEKLNARFCSAKIKAVFGLLPVLSMLPKY 65  
DB 15 RYVVERPFSHPVLOERLHVMDKVTESIGDKKQAFYTPKKIRNIIIMFLPTKMLPAY 74  
QY 66 KIKDYIIPDLGLSGSIOVPQGMARFALLANPAVNGLYSFFPLTYFFLGVHOMVP 125  
DB 75 KFEKRYVLDLVSGISTGVLOLPQGLAFAMLAAPVFGLYSFFVIMVCEFGTSRHISI 134  
QY 126 GFPAVISILVGNICLOLAPES-KFOVFNNATNESYVDTAAMEARLHSATLACTLTAIO 184  
DB 135 GFPAVISILMIGVAVRALVPDDIVIRGVNATN---GTARDALRYKAMASTLLSGITIO 190  
QY 185 MGLGFMQGFVAIYLSSEFIRGEFTAGLIQILISLVKYLFGLTIPSYTGPDSIVETFDI 244  
DB 191 FCLGRCRGFAIYLTLEPLRGFTTAAAVHFTSMKLYLFGVKTARGSIFSVYSTAV 250  
QY 245 CKNLPHTIASL-----IFALISGAFVLVKELNARMKIRFPPTKEMIVYVATAISG 299  
DB 251 LQNVKMLNVCISLGVLMFGLLG---KEFNRFKEKPPAPLPLEFFAVVMGTGISA 305  
QY 300 GCKMPKRYHQMOIVGEIQRGFPPTVPSPVVSQMKDMGTAFSLAIVSYVNLAMGRTLANKH 359  
DB 306 GPNLHESIVDVVGTLPGLPRLPAPNPTSLRHLVYVDALAIAYGFSVTISAKTLANKH 365  
QY 360 GYVDVDSNQMIALGCSNFFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMITM 419  
DB 366 GYVDVDSNQMIALGCSNFFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMITM 425  
QY 420 LVLGILYLPKRSVIGALIAVNLKNSLKQLDPTYLMRKSILDCIIVVVSFLSFLSLP 479  
DB 426 LATGFLFESLPOAVLSAIVNLKGMFQSDLPFMRKSKIEPTIMLTTEVSSFLGLD 485  
QY 480 YGVANAVFSVLVVFQFRNGYALAOVMDIYVNPRTNRADIDIGIKIITCSPLY 539  
DB 486 YGLTAVIATLTYIKTQSPSYKVLGQDPDVIDIDDAEVEKELFGIKIIFQINAPLY 545  
QY 540 FANSEIFROKVIATY-----VSLOLODFENAPPTDPNNQTPANGTSVSYI 587  
DB 546 YANSDLYSALKRKTKGVNAPALIMGARAKMRKAYKEVGNA-----NVANATVVKYDAE 598  
QY 588 TFSPPSSSPAQSE-----PPASAEAPGEPMDLASVPPVTFPHLLIDMSGVSEVDLMGI 642  
DB 599 VDGEMATPEEEDDEVKPPPIVIKTT-PPETLORELPOGEMVHTVILDFTOVNEFVDSGV 657  
QY 643 KALAKLSLTYGKIGKVVLYNINHOVYNDISHGVFEEDSGLECKKIVFSPSHDAVYFAQ 700  
DB 658 KTLGAIYKEYGDVGIYVLACGSPQVYNDLTRNNEFENPALK-ELLFHSIDAVYLSGQ 714

RESULT 6  
ID 09R155 PRELIMINARY; PRT: 780 AA.  
AC 09R155;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PENDRIN.  
CN SLC26A4 OR PDS.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_Taxid=10090.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/C;  
RA MEDLINE=99380587; PubMed=10449762;  
RX Everett L.A., Morsli H., Wu D.K., Green E.D.;  
RT "Expression Pattern of the Mouse Ortholog of the Pendred Syndrome Gene  
(Pds) Suggests a Key Role for Pendrin in the Inner Ear."  
RL Proc. Natl. Acad. Sci. U.S.A. 96:9727-9732(1999).  
DR EMBL; AF167411; AAD51617.1; -  
DR MGD; MGI:1346029; SLC26A4.  
DR InterPro; IPR002645; STAS.  
DR InterPro; IPR001902; Sulfate\_transp.  
DR Pfam; PF01740; STAS; 1.  
DR Pfam; PF00916; Sulfate\_transp; 1.  
SQ SEQUENCE 780 AA; 85686 MM; 4988B48F057BF38 CRC64;

Query Match 30.6%; Score 1185; DB 11; Length 780;  
Best Local Similarity 35.1%; Pred. No. 3.8e-66;  
Matches 272; Conservative 147; Mismatches 297; Indels 60; Gaps 13;

QY 8 YVVDRAVYSLTLEDFDEFEKDRTPYGEKLNARF-----RCSAKIKAVFGLLPVLSMLP 63  
DB 20 YVVSPPVYSELAFQ---QQERRRLPERRTLRSLARSCSKRAFGVVKTLPLDMLP 76  
QY 64 KYKIDYIIPDLGLSGSIOVPQGMARFALLANPAVNGLYSFFPLTYFFLGVHOM 123  
DB 77 KIRAKEMILSDIISGVSTGLVGTLDGMAYALLAAPPVFGLSAFPLITTFVETSNHI 136  
QY 124 VPGFAVISILVGNICLOLAPESKFOV---FNMATNESYVDTAAMEARLHSATLACT 180  
DB 137 SVGPFPVYSLWVGSVLSAMAPDDHFLVPSNGSALNSTLDTGTDARVLLASTLTLV 196  
QY 181 AIIQMGCLPMQGFVAIYLSSEFIRGEFTAGLIQILISLVKYLFGLTIPSTYTPGSIYFT 240  
DB 197 GIQLVFGGLDIGFIRYADLPVGGFTTAAAFQVLSQILVNLVSTKNNGILSIYIT 256  
QY 241 FIDCKNLPHNINIASLIFALISGAFVLVKELNARMKIRFPPTKEMIVYVATAISG 300  
DB 257 LIETIQNTGDMIDAFIAGLLITIVCMKAVKELNDFKRRIRVPIPIEIVYITIAISYG 316  
QY 301 CKMPKRYHQMOIVGEIQRGFPPTVPSPVVSQMKDMGTAFSLAIVSYVNLAMGRTLANKH 360  
DB 317 ANLEKNYAGIYKSLPSGFLPVLPSVGLFSDMLAASSIAVNAALAVSGKYATFKHD 376  
QY 361 AIIQMGCLPMQGFVAIYLSSEFIRGEFTAGLIQILISLVKYLFGLTIPSTYTPGSIYFT 240  
DB 197 GIQLVFGGLDIGFIRYADLPVGGFTTAAAFQVLSQILVNLVSTKNNGILSIYIT 256  
QY 241 FIDCKNLPHNINIASLIFALISGAFVLVKELNARMKIRFPPTKEMIVYVATAISG 300  
DB 257 LIETIQNTGDMIDAFIAGLLITIVCMKAVKELNDFKRRIRVPIPIEIVYITIAISYG 316  
QY 301 CKMPKRYHQMOIVGEIQRGFPPTVPSPVVSQMKDMGTAFSLAIVSYVNLAMGRTLANKH 360  
DB 317 ANLEKNYAGIYKSLPSGFLPVLPSVGLFSDMLAASSIAVNAALAVSGKYATFKHD 376  
QY 361 YDVSNDQMIALGCSNFFGSFFKIHVICALSVTLAVDAGAGKSOVASLCSLVVMITM 420  
DB 377 YVIDNQEFIAGISNVFSGFSCFVATTAASRTAVQESTGKIQVAGLISAVIYMAIV 436  
QY 421 VLGITLYLPKRSVIGALIAVNLKNSLKQLDPTYLMRKSILDCIIVVVSFLSFLSLP 479  
DB 426 LATGFLFESLPOAVLSAIVNLKGMFQSDLPFMRKSKIEPTIMLTTEVSSFLGLD 485  
QY 480 YGVANAVFSVLVVFQFRNGYALAOVMDIYVNPRTNRADIDIGIKIITCSPLY 539  
DB 486 YGLTAVIATLTYIKTQSPSYKVLGQDPDVIDIDDAEVEKELFGIKIIFQINAPLY 545  
QY 540 FANSEIFROKVIATY-----VSLOLODFENAPPTDPNNQTPANGTSVSYI 587  
DB 546 YANSDLYSALKRKTKGVNAPALIMGARAKMRKAYKEVGNA-----NVANATVVKYDAE 598  
QY 588 TFSPPSSSPAQSE-----PPASAEAPGEPMDLASVPPVTFPHLLIDMSGVSEVDLMGI 642  
DB 599 VDGEMATPEEEDDEVKPPPIVIKTT-PPETLORELPOGEMVHTVILDFTOVNEFVDSGV 657  
QY 643 KALAKLSLTYGKIGKVVLYNINHOVYNDISHGVFEEDSGLECKKIVFSPSHDAVYFAQ 700  
DB 614 AFEDE-----DVEPEELNPTKEIEIQVDSNSELVYKVVNPKVPIHSLVLDGCAVSFLD 669



[illegible]

DB	435	IVALGLKEPLQKSVLAANVIANLKGMMQVCDVPRILKQKKTDAVIVTFWICHSIIIGL	494
QY	479	PGYVAAGVAFSVLVVVFQTOFRNGYALAOVMDTDIYVNPCKTYNRAQDIQIKITTYCSPL	538
DB	495	DGLLAGLLEGLLVTVLRQFPFSNMGSLVSPSTDIYKSTIYHKNLEBEGVKILRFSSPI	554
QY	539	YFANSEIRPROKIAYKT-----SLQELQDFENAPRPDPNNQTPANGTSVS	585
DB	555	FPGVNDGFR-KCVKSTVGFDARVYNKRLKRLARRLOKLIKQOLRATNGIISDGGSSNN	613
QY	586	YTFESPSSPQSPQSPASAEAPGE-----PSDMLASVP-----PRVPTHTLLDMGSVSF	636
DB	614	--AFEPDE-----DVEPELDLPITKEIETQVDMNSELYKKVNPVPHSLYLDGCAVSF	667
QY	637	VDLMGSIKALAKLSSTYGGIKGVKFLVNIHAQYNDISHGVEFSDGLECKHVPSIDHAY	696
DB	668	LIVGVRSLRMIVKREFGRIDVNVYFALLQDDVLEKMEQCGFFDD-NIKKDFEFLVHDAI	726
QY	697	LFAQANAR 704	
DB	727	LYLQNOAK 734	
RESULT 8			
Q9WVC8	Q9WVC8	PRELIMINARY; PRT: 757 AA.	
AC	Q9WVC8;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	DOWN-REGULATED IN ADENOMA PROTEIN.		
GN	SLC26A3 OR DRA.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N; TISSUE=COLON;		
RA	Melvin J.E., Park K., Richardson L.A., Schultheis P.J., Shull G.E.;		
RT	"Mouse Down-Regulated in Adenoma (DRA) is an Intestinal Cl-/HCO3-		
RT	Exchanger and is Upregulated in Colon of Mice Lacking the NHE3 Na+/H+		
RT	Exchanger."		
RL	J. Biol. Chem. 0:0-0(1999).		
DR	EMBL; AF136751; AAD42784.1; -		
DR	MGD; MGI:101781; SLC26a3.		
DR	InterPro; IPR002645; STAS.		
DR	InterPro; IPR001902; Sulfate_transp.		
DR	Pfam; PF01740; STAS; 1.		
DR	Pfam; PF00916; Sulfate_transp; 1.		
DR	PROSITE; PS01130; SULFATE_TRANS; UNKNOWN.1.		
SO	SEQUENCE 757 AA; 83589 MW; 32B1AC648BE74A07 CNC64;		
Query Match 29.7%; Score 1147.5; DB 11; Length 757;			
Best Local Similarity 34.7%; Pred. No. 8.2e-64;			
Matches 254; Conservative 149; Mismatches 261; Indels 67; Gaps 10;			
DB	7	RIVVDRAATSLTLPDDEFEKDRITPVG-EKLRNAFRCSSAKIKAVGCLPVLSMLPKY	65
QY	8	YVVAVRPVYSTTFEEFEKTKRHNRKTFDLHLKGCSSSOCAKRIALSTEPFIASMLPAY	67
DB	66	KIKDVTIPPLTGLSGSGSIOVPOGMAPFALIANLPAVNGLYSSFFPLLYTFPLGVHWP	125
QY	68	KIKEMILSDIVSGISTGLVAVLQGLAFALLVNIIPAYGLKAAFFVYITFFLGTSRHLSV	127
DB	126	GTFEAVISLVNICQLA--PESKQOVFNNAF-NESYVDTAAMEERLHVSATLACRAI	182
QY	128	GPEFVLSMMVGVVTRVSDPNAASELSSSTENDSFIE-----EKVVAASVTVLSGI	181
DB	183	IOMGCGFMQFGVVAIYLSSEFTRGFWMTAAGLOILISLVKTYIFGLTIPSTYTGPGSIVTFPI	242
QY	182	IOLLGLGVLDGVGVVYILSSLSISGFTTAAAHIVLVSOLKFMQLQPLVPPVSDPFSIFKYLE	241

QY	243	DICKNLPHTNLSIAFLIALISAFVLVKEENARXMHIRRPIDREMIYVVAVATASGCK	302
Db	242	SVFTOIQTNTADLTVSYIIIVVFVEFIEIQRKSLPWPPIELMTIYATIGVSGCN	301
QY	303	MPKKYHMOIVGEIQRGFPVPVPSQWKMDIGTAFSLATISVYINIAMRTLANKHGD	362
Db	302	FEDRFGVAVVANNMSIGFOPPIITPSVEVEFQDITGDSFGIAIGFAVAFVSASVSLKYDP	361
QY	363	VDSNEMIALICGSPFGSFEFFIHHVICALSVTLAVDAGGKSOYASLCVSLVYMTMLV	422
Db	362	IDGNELIALVSNLEFGAFGFGAGSTRLSKSGVOSTGKTQVAGLNAVILYITVAI	421
QY	423	GIYIYPLKRSVGLALVANNLSLKQITDPYLLTRSKSLDCCIMVVSFLLSPFYGV	482
Db	422	GFLLOPLQKSVIALAALGNTLKGMLMFOAEIGRELMKKDKYDCLIMTFEIAVIGLGL	481
QY	483	AVGVAFSVLVVFPQTFPENGALAAQWMDTYVNPKTYNNRQDIOGKITTYCSPLYFAN	542
Db	482	AASVAFQLLTITVFRTQFKCSTLANVGSNSNKKNNYAEVEGKIFRCPSPIYFAN	541
QY	543	SEIFPROKVI-----AKTVSIOELOQ-----DEENAPPTD	571
Db	542	IGGFQKTLIDAVGVSPURLKRMKALKIKIKLQKGLQMTPKGICTSDGRKDSDEL	601
QY	572	PRNN-----QTANGTSVSY-ITFSPDSSSPQOSEPPASAEAPGEPSSMLASVPEYTFHT	626
Db	602	DNNQLELDQPIINTDTCDFIDMNGDLP-----LWITTIPIKSLHS	641
QY	627	LILDSGVSFPVLDLGIKMLKALSTSYGKIVGVFLVNIHAQVYNDISHGCVFEDGSLECK	686
Db	642	LILDSAVSFLDVSMSRGLRPTITQIEFIRIKVDYIVGTDDQFIDKLARCFPD-ETVDS	700
QY	687	HVFPSIHDAVL 697	
Db	701	IFELTTHDAIL 711	
RESULT	9		
Q9BEG8	Q9BEG8	PRELIMINARY;	PTT; 734 AA.
AC	Q9BEG8:		
DT	01-JUN-2001 (Tremblrel, 17, Created)		
DT	01-JUN-2001 (Tremblrel, 17, Last sequence update)		
DT	01-JUN-2001 (Tremblrel, 17, Last annotation update)		
DE	DTD SULFATE TRANSPORTER.		
OS	Bos taurus (Bovine).		
CC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
NC	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Baumgartner B.G., Kriegesmann B., Brenig B.;		
RT	"Cloning and characterization of the bovine DTD sulfate transporter		
RL	(dbpdsr) gene."		
SR	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ223615; CAB69640.1; -		
Q9	SEQUENCE /34 AA; 81540 MW; D5224A27EA5D691E CRC64;		

```

Query Match      26.3%; Score 1016.5; DB: 6; Length 734;
Best Local Similarity 32.8%; Pred. NO. 1.2e-55;
Matches 231; Conservative 137; Mismatches 286; Indels 51; Gaps 9;

Oy 25 EKKDTP--VGEKLRNFRCSAAIKAAVPEGLLVLSLWPKRYTKIDYIIPDLGLSG 82
    |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 61 EKSDNNFKKFKVKKLEKSCSSTAKNTIFGFLPVLQWLPKPYDLKKNILGVMSSLLIG 120
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||

Oy 83 SLOVPGMAFALNALPVPANGLYSFPFLITFEFLGVQWMPGRFAVISLVGNICQL 142
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||
Db 121 ILVPOSTAVSLLAQOEPLRYGLTYSFMSLIFYILGTSTSHIVSGIFGILCLMIGEVNDE 180
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::||

```

0Y	143	APESKFOYFNNAATNSYV-----DTAMEERLHVSAITLACLAIIOMGJFMPGOFEVA	156
Db	181	LYIACDTVHAASNNSSLVNOISDKTCDRCRYAIIIVGSTVFYAGVIOVANGFQGVFS	210
0Y	197	ILYSESPTRGEMTAAGLOILILSVLYIFGLTIPSTGPGSVTFEIDICNKLPHNTIASL	256
Db	241	YLSOALLGRTVGASFTILTSQYKLLGLISLPSRAGVGSILTTWLVHFNIRKTNICDL	300
0Y	257	IFALISGAFVLVVELNARRYMKRTREPIPEMIYVVAATAISGGCKMPKRYHMOYGEIQ	316
Db	301	ITSLCLCLIVLLPTKELNERFKSKLAPIPVELFVIVATLASHGKLNERYGTSIAGHIP	360
0Y	317	RGEFPPVSPVVSQMKMDTGTFASLAIYSVYNILAMGRTLANKHEDYDSNOEMTALCSN	376
Db	361	TGFMPKPAKDMNLLIPRAVADIAIATIGFALTVLSSEMFARKHGYTKANDEMAITFCN	420
0Y	377	FEGSEFFKIHVICALSVLAVDAGAGSKQVNSLCSVSLVMTVMYTLVLSIYLPLEKSVLGA	436
Db	421	IIPSEFHCFITTSALAKLTVKESNGCGOQVSGWMTLVLLVLTIVAPLPEFSLOKSVLGV	480
0Y	437	LIAVNLKNSKQLTDPYILMRKSKLDCCIVVVSFLSEFISLPYGAVGAFSVLVVVFQ	496
Db	481	ITIVLRCALCKRFDLPOMMRISHMDDYIMFVMTLSALSTEIGLLTGOCFSMFCYILR	540
0Y	497	TQFRNGYALAOYMDIYVNRKTYNNRQDIOGIIITTCSPLYANSEIROKVIATYS	556
Db	541	TQKPRASLLGLVESEVESASAKNLQAKSGIIFRRVAPLVVYVNEFKSVLYKKTTLN	600
0Y	557	LOELQODENAPRPTDNNQNTPANGTSVSYTTFSPDSSPPOSEPPASAPAPGESPDLA	616
Db	601	-----PVLKRAAQKKAKKRKRKRETYIP---SGIQDE--VSYQLSDHP-----	638
0Y	617	SVPEPVTEHTLILDSGVSFVLDLGIKALAKLSTSYGKIGVKEFLVNIHAQVINDISHGG	676
Db	639	-----LEEHITVICSALQIFDFTAGIHITLKEVRDYDEAVGIOVLACNPSVRDSLARGE	693
0Y	677	VFEDOSLECKH-----VFPSITHDAVLPAO-----ANARDYVPGHNF	712
Db	694	Y-----CKKDEENLFTYSIYEMATPEASQONOKERHINPGPNF	731

```

RESULT      10
O9GUY3
ID      O9GUY3      PRELIMINARY;      PRT;      734 AA.
AC      O9GUY3;
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      SULFATE TRANSPORTER.
GN      ST.
OS      Ovis aries (Sheep).
OC      Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Caprinae; Ovis.
CX      NCBI_TaxID=9940;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Kriegermann B., Baungartner B.G., Deppe A., Brenig B.,
RT      "Sequence of the ovine sulfate transporter gene.";
RL      Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Y18558; CAC20729.1; -.
DR      InterPro: IPR002645; STAS.
DR      InterPro: IPR001902; Sulfate_transp.
DR      Pfam: PF001740; STAS: 1.
DR      Pfam: PF00916; Sulfate_transp. 1.
DR      PROSITE: PS01130; SULFATE_TRANSP. 1.
SQ      SEQUENCE 734 AA: 81439 MW: 3535EADFB4F2F3B36 CRC64;

```

Query Match	25.98;	Score 1001.5;	DB 6;	Length 734;
Best Local Similarity	33.28;	Pred. No. 1.1e-54;		
Matches 229;	Conservative 122;	Mismatches 281;	Indels 47;	Gaps 8;

Oy		25	EKKRRTPV--VGEELRNAPRSSKIKIAVEGGLPVLMSLPKRRIKOYIIIPDLGGISGC	82
Db		61	:     :     :     :                 :   :   :     :	
Oy		83	SIOVPQGAFFALLNIPRAVNGLSSFPFLLYTFEPLGCGHOMVGCTEAVISILVGNICLOL	142
Db		121	:     :     :     :                     :   :   :     :	
Oy		143	APESEKFOVFNNATNESTV-----DTAMEAERLIHSATLACTLTAILOMGLGFMOEGFA	196
Db		181	LYIAGYDVTVAHASNESSLVNMOSNOTCDRSCLATTVGSTVTFVAGAYQVAMGFEQQGVFS	240
Oy		197	IYLSESFIRGMTAAQIOIILSVLKIFGLTIPYTPRGSIIVFPIFDICNLPHNTIASL	256
Db		241	VYLSLDALLGCVTASFTITTSQVKYLLIGSLSPRSGGVSLITTWIHIFPNHKTNICDL	300
Oy		257	IFALISGAFVLVELNARVYMKTRFPIPTEMIYVVATAISGCCPKPRYHQIYEIQ	316
Db		301	ITSLLCILLVLTPELENERFRSKAKAIPELPPVVAATLASHGKISEKYSIACHIP	360
Oy		317	RGPPTPVSPVVSQMKDMIGTAFSLAISVYINLAMGRTLANKHGVDYSNOEMIALGCSN	376
Db		361	TGFMPKPARDWMLIPRAVDAIDAIIGCFATVSLSEMFAKHGYTKANQEMVAIGFCN	420
Oy		377	FFGSFEPKIHVLCALSTYLANDAGSGSQVASLVCSLVIMTLVLCITYLPRLPSYLG	436
Db		421	IIPFESHFTTSAALACTLVKESEGCCOTGVSWTALVLLVLLVLAIPFESLOSXYLVG	480
Oy		437	LIAVNLKMSKQLNDPYVIAMKSKLDDCIWVSFLSFPFLSPGVAVGAFTLVVPO	496
Db		481	ITVYNLRKALCKFKFDLPQMHRISMEDVIMVMTLSSALISTEIGLTGVCFSMFVILR	540
Oy		497	TQFNRYALAOVMDTDIVYNNKYTNRRADODLOGIKIITYCESPLYFANSEIPROKVIATVS	556
Db		541	TQKRKASLLGLVEDESEVFESMSAKKNQANSKGIIEFPVAPLVYNNKEYKSYLVKKTLN	600
Oy		557	LOELQODEFENAPPTDPNNNOTPRANGTSVSYTESPDSSSPAPOSEPPASAABGPSOMLA	616
Db		601	-----PVLVYAQAQRKAKKRIKRETV---TlSGIODE--VSVOLSDP-----	638
Oy		617	SVPPFVTFHTLILMSGSPFDLMGIALKLKLSSTGYKIGVYKVLVINHAQVYNDISHGG	676
Db		639	----LEFHITVIDCSALOFPDTRAGHTLLEVARDYEAIGIYVLLAACNSVSRDSLARGE	693
Oy		677	VFEDEGSLECKH-----VEPSIHDAVLPQAQ 700	
Db		694	Y-----CKKDEENLLEYSYEAMETPAE 715	
RESULT	11			
O9BGH1				
AC	O9BGH1	PRELIMINARY;	PRT; 463 AA.	
PT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	DOWN-REGULATED IN ADENOMA DRA (FRAGMENT).			
GN	SLOC26A3.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID=9986;			
RA	SEQUENCE FROM N.A.			
RC	TISSUE=INTESTINAL MUCOSA;			
RA	Jacob P., Rossmann H., Lamprecht G., Kretz A., Neff C., Lin-Wu F.,			
RT	Gregor M., Seidler U.;			
RT	*Characterization of rat and rabbit brush border membrane anion			
RL	exchange and DRA expression in rabbit, rat, and human duodenum.*;			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF314819; AAK00897.1; -.			
FT	NON_TER 1 1			
FT	NON_TER 463 463			

[illegible]

Db 2 NIPPAVGLYAAAEFPVITYFFELGTSHRISVGPPLVSMVGVVTVRVASGD--TSPALS 58  
 Qy 157 ESYVDTAAMEAEERLHVSAITLACTTAIOMGLGFMQGFVAIYLSSEFIRGFMTAGLOIL 216  
 Db 59 SSSAENDSIEERKVVAAVSVTLGSLIQOLLGVLQIGFVITYLSGSLFTTAALINVL 118  
 Qy 217 ISLVKTYFGLTIPSTYGPESIVFTFDICKNLPHTNIALSLFALISGAFLVVKELNARY 276  
 Db 119 VSQLFMQLQITVPASDPSIFKVLSEVSQIQKTNIDLVTSVILLVVFAVKEINQRY 178  
 Qy 277 MHKIFPIPTFEMIVVVVVAISGCKMPKKYHMOIVGELQREPPRPVSVSOMKDMCT 336  
 Db 179 RSKLVPPIELIMTVIATIGISYGCNFEORFEGVAVGNNSLGFQPIPTPSVEFDDTIGD 238  
 Qy 337 AFSLAIVSVIIMAMGRTLANKGYDVDSNQEMIALGCSNFGSFEKIHVICALSVTLA 396  
 Db 239 CGCIAIVGVAFAVSVAIVSLKHYPIDGHQELIALGVSNITGAFKGNASTALSRSV 298  
 Qy 397 VDGAGCKSOVASLCVSLVVMITMLVGLIYLPKPSVGCALIAVLKNSLKOLTPDYIUM 456  
 Db 299 QESXGKTOVAGLISAVIYLIYVAIGFLQPLQKSVLAALAGLKGMLQFAEIGRLM 358  
 Qy 457 RSKKIDCCIMVVSFLSPFLSPYGVAVGAVSVLVVVFQRFNGYALAOVMDTDIYVN 516  
 Db 359 KDKDKDCLIMTFTFAIYLGIGLGLASVAFOLITVFRTOFPKCSITLANGRSNITKN 418  
 Qy 517 PKTYNRADIDIGIKITTYCSPLYFANSEIFROKVI 551  
 Db 419 KKNVADYVEPEGVKIFRCSPPIYFANIGFEKOLI 453

## RESULT 13

Q9H2B4 PRELIMINARY: PRT; 701 AA.  
 Q9H2B4  
 AC 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SUIFAIE/ANION TRANSPORTER SAT-1 PROTEIN.  
 GN SLC26A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE=20541715; PubMed=11087667;  
 RA Lohli H., Kujala M., Kerkela E., Saarialho-kere U., Kestila M.,  
 Kere J.;  
 RT "Mapping of five New Putative Anion Transporter Genes in Human and  
 RT Characterization of SLC26A6, A Candidate Gene for Pancreatic Anion  
 RT Exchanger.";  
 RL Genomics 70:102-112(2000).  
 DR EMBL: AF297659; AAC22075.1; -  
 DR InterPro: IPR003880; Phosphopant\_attach.  
 DR InterPro: IPR002645; STRAS.  
 DR Pfam: PF01740; Sulfate\_transp.  
 DR Pfam: PF00916; Sulfate\_transp. 1.  
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN\_1.  
 SO SEQUENCE 701 AA; 74991 MW; F37DC87F6A036676 CRC64;

## Query Match

22.7% Score 878; DB 4; Length 701;

Best Local Similarity 29.6%; Pred. No. 5.2e-47; Indels 46; Gaps 8;

Matches 209; Conservative 146; Mismatches 304;

Qy 29 RTYPVGEKLRNFR-----CSSAKIKAVFGLPLVSLPKKIKYIIPDLGLGS 80  
 Db 19 RORPARGLREMLKARLWCSGSCSVLCVRAVLQODLLPATRMLRGYRPREVLAGDVMSGLV 78  
 Qy 81 GGSIOVPGQMAFALLANLPVANGLYSFFPLLTTFYFLGCVHOMVPGTFAVISILVGNIC- 139

Db 79 IGIIIVPOAIAVSLLAGQIPYISLYTSFANLIYFLMGSRHVSIGFISLCLAMGOVVD 138  
 Qy 140 --LOLA--PESEFOYFNNAFTNESYVDTAAM-----EAEELHSAITLACTTAIOMGLGFM 190  
 Db 139 RELQLAGFDPSSDGLQPGANSSSTLNGSAMLDCGDCCAIRATATLTMTGLYOLMGVL 198  
 Qy 191 QGFVAIYLSSEFIRGFMTAGLOILISVLYKTYFGLTIPSTYGPESIVFTFDICKNLPH 250  
 Db 199 RLGFVSATLSQPLDGFPMAGASVTILTSLQKLGLVRIPIRQGPQMVVLTMTSLLRGAGQ 258  
 Qy 251 TNIASLITRALISGAFLVVKELNARYMHKIRPITFEMIVVVVVAISGCKMPKKYHMO 310  
 Db 259 ANVCQVVTSTVCLAVLAAKELSDRYRRRLVPLTELLIYVATVSHFGQLHRRFGSS 318  
 Qy 311 IYGEIORGPPRPVSVSOMKDMIGTAPSLAVSVIIVNAMGRTLANKHGYDVDSNQEMI 370  
 Db 319 VAGDIPTGFMPOVEPRRLMQRVALDAVALVAAFSISLEMPARSHGVSYRANQELL 378  
 Qy 371 ALGCSNFGSFEKIHVICALSVTLAVDAGCKSOVASLCVSLVVMITMLVGLIYLP 430  
 Db 379 AVGCCNVLPAFLHCPATSAALAKSLVKTATGCRTOSSVSATVLLVLALAPLEHDLQ 438  
 Qy 431 KSVLCALLAVNLKNSLKOLTPDYIYLMRKSILDCCLMVVSFLSFLSLPYGAVAVASV 490  
 Db 439 RSVLACIVYSLRGALRKWGFPRRLMRSPADALWAGAACMLVSTEGAGLAVILSL 498  
 Qy 491 LVVVFOTOFNGYALAOVMDTDIYVNPNTYNRADIDIGIKITTYCSPLYFANSEIFROK 550  
 Db 499 LSLAGRTQRPRTALLARIGDTAFYEDATEFEBGLVEPGVRYRFRGGLPYANKOEFLOSL 558  
 Qy 551 IAKTVSLOELQDFENAPPTDENNOTPANGTSVSYITFSPDSSPAQSEPPASAEADGE 610  
 Db 559 ----YSLGILDAGCMARRKEGSETEGEG-----GPAQGEDLDGP 595  
 Qy 611 PSDMLASVPPFTYFHTLLIDMSGVSFVDMGLKALAKISTYTGKIGVAVFLNIAHYVN 670  
 Db 596 VSTRALVPAAGFHTVYIDCAPLFLDLAAGVSTLDLRDGLAGISILLACSSPPVRD 655  
 Qy 671 DISHGVEFDG---SLECKHVPSTHDAVLEFAQANARV--TPGH 710  
 Db 656 ILSRGLFGEGRGDPAEEBQFLSVHDVNOTRAHRELENDAH 700

## RESULT 14

Q9VVM6 PRELIMINARY: PRT; 742 AA.

Q9VVM6  
 AC 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CG5485 PROTEIN.  
 GN CG5485.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,  
 Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,  
 Botkova D., Botchan M.A., Bouck J., Brokstein P., Brotter P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,



QY	138	ICQLABESKFOVENNATNESYVDVTAMAEARLHVSAATLACLTALITOMLOJEMOGEFVAI	197
Db	114	ARLRALRD--IPISNSSDIMPVSYPPLGEYVDLPVTFSAITLLGVQOIMGLIRLGFLLT	171
QY	198	YLSSEIFRGMTAAGLOILISVLKTYIFGLTIPSTYBPBSIVTFEDICKNLPHTNASLI	257
Db	172	YLSDSLVSGFTTGAAVHVFSTOLNKVFGILPRHNEGIMIVRMVROMIMSLGSVNFVALG	231
QY	258	FALDSGAFVLYVELNRYNKHK1-----REPITEMIIVVYVATISGGCKMKPKYMOIY	312
Db	232	ISI-----FGLIFDLGTTYINPIYKKFSPLPPELLELIVIFGVISMIRNDAENHVKTIV	287
QY	313	GEIQRGFPPTVSPVYSQWKDMIGTAFSLAIVSYVINLANGRTLANKGYDVDSNOEMIAL	372
Db	288	YEIRGRGPEPSIPLRNLPALLSDAIPRAIVVCMYFMSMGKLEFAKHKHKYKTDATGCLAI	347
QY	373	GCSNFPSFKRINHICALSVTLAVDAGAGSQVA5LCSLVYVMITMLVLYGILVLPPLRS	432
Db	348	GIASALSSFPVYPVGSILSRSSVCEMSGANTLOYITFSSFLLYVILLIGFPLEPLPMC	407
QY	433	VGLALIVLNKNSLKQLTDEYUUMRRKSKLDIMVVSFLSSPFLSLPYGVAAGVAFSVLY	492
Db	408	ILACTIVLSLKSLEMOVKELPRILYRISKYDFALMTLVACLSITFEYDVYTGVLVSLAFSLY	467
QY	493	VVFOTQFRNGTALAQVMDTDIYVNPXTYNAQDI-OGIKITTYCSLYFANSEIFROKYI	551
Db	468	LYLRQOMP-----SFSTYHDEK--TPR-----QWVPENKVIYFAGSLHEANVTAF-----	511
QY	552	AKTVSLQELQODDFENAPPTDPNNQIPRANGTSYITFSPDSSSPROSPRASALEARCPEP	611
Db	512	-----ODDM-----GFA	518
QY	612	SDMLASVPPEVTEHTLLIDMSGV5FYDLMGIKALASSTYGRIGVKEVLVNIHQVYND	671
Db	519	IGKLPREDPLIDETIILLDASSVAFIDIMGVDAIRDVFKDALKIGVHVEYCGLPREDVLSY	578
QY	672	ISHGCVFEDDSLECKNHVPSIHNHVALEFAQANADY	706
Db	579	LSNDENFL-SVWPPSTFSPSI-DSCLSIFVHQO5Y	611

Search completed: April 26, 2002, 09:06:43  
Job time: 134 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:04:39 ; Search time 12.77 Seconds  
(without alignments)  
2161.990 Million cell updates/sec

Title: US-09-749-589-2

Perfect score: 3869  
Sequence: 1 MSQPRPRRYVDRAVSLTLE.....WDLQEMFGSMFAETLTAL 753

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1190	30.8	780	1	PEND_HUMAN
2	1120	28.9	764	1	DRA_HUMAN
3	974.5	25.2	739	1	DTD_RAT
4	972	25.1	739	1	DTD_HUMAN
5	954.5	24.7	703	1	DTD_MOUSE
6	853	22.0	703	1	SAT1_RAT
7	588	15.2	758	1	YLDM_CAEEL
8	518	13.4	662	1	SUT2_STYHA
9	516	13.3	667	1	SUT1_STYHA
10	496	12.8	644	1	SUT3_STYHA
11	464.5	12.0	877	1	SULH_SCHPO
12	437	11.3	754	1	SULX_YEAST
13	405	10.5	893	1	SULZ_YEAST
14	377	9.7	859	1	SULL_YEAST
15	375.5	9.7	485	1	NO70_SORBN
16	357	9.2	788	1	CY14_NEDCR
17	230.5	6.0	550	1	YCHM_ECOLI
18	181	4.7	434	1	YBAR_BACSU
19	135	3.5	1036	1	YG35_YEAST
20	132	3.4	442	1	CITN_LACIA
21	125	3.2	526	1	MYIN_TREPA
22	124	3.2	654	1	PSTA_MYCGE
23	121	3.1	958	1	VA7B_SCHPO
24	121	3.1	1276	1	MR1_MOUSE
25	119.5	3.1	521	1	ND2C_SYNF7
26	116.5	3.0	445	1	YIEG_ECOLI
27	116	3.0	568	1	PTLB_STRMU
28	115	3.0	633	1	Y147_HAEIN
29	114.5	3.0	400	1	NHAA_HAEIN
30	113.5	2.9	470	1	YICO_ECOLI
31	113.5	2.9	533	1	MYIN_SYNF3
32	113.5	2.9	636	1	CTR4_HUMAN
33	113.5	2.9	1232	1	B3A3_HUMAN

34	112.5	2.9	471	1	NORM_CAUCR	P58163 caulobacter
35	111.5	2.9	525	1	WMT1_HUMAN	P54219 homo sapien
36	111.5	2.9	545	1	N05M_ALBCO	P48918 albinaria c
37	110.5	2.9	394	1	BERNE_ACICA	P07775 actinobact
38	110.5	2.9	715	1	LCNC_LACIA	Q00564 lactococc
39	110.5	2.9	763	1	RCGT_YEAST	Q12300 saccharomyc
40	109.5	2.8	434	1	YHFT_ECOLI	P45546 escherichia
41	109.5	2.8	744	1	N05C_GERJA	P51100 gerbera jam
42	109	2.8	452	1	NORM_BACSU	O31855 bacillus su
43	108.5	2.8	438	1	SECY_METVA	P28541 methanococ
44	108	2.8	561	1	BIO5_YEAST	P53744 saccharomyc
45	107.5	2.8	699	1	N05C_DIGGR	Q32131 digitallis g

## ALIGNMENTS

RESULT	ID	PEND_HUMAN	STANDARD	PRT	780 AA.
1	AC	043511; 043170;			
	DT	15-JUL-1999 (Rel. 38, Created)			
	DT	15-JUL-1999 (Rel. 38, Last sequence update)			
	DT	20-AUG-2001 (Rel. 40, Last annotation update)			
	DE	PENDRIN (SODIUM-INDEPENDENT CHLORIDE/IODIDE TRANSPORTER).			
	GN	PDS.			
	OS	Homo sapiens (Human).			
	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
	OX	NCBI_TaxID=9606;			
	RN	[1]			
	RP	SEQUENCE FROM N.A., AND VARIANT PDS CYS-667.			
	RC	TISSUE=Thyroid;			
	RX	MEDLINE=96061089; PubMed=9398842;			
	RA	Everett L.A., Glaser B., Beck J.C., Idol J.R., Buchs A., Heyman M.,			
	RA	Adavi F., Hazani E., Nassir E., Baxevanis A.D., Sheffield V.C.,			
	RA	Green E.D.;			
	RT	"Pendred syndrome is caused by mutations in a putative sulphate			
	RT	transporter gene (PDS).";			
	RL	Nat. Genet. 17:411-422(1997).			
	RN	[2]			
	RP	SEQUENCE OF 336-780 FROM N.A.			
	RA	Smith A., Johnson D., Harmon G.;			
	RL	Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.			
	RN	[3]			
	RP	FUNCTION.			
	RX	MEDLINE=99206617; PubMed=10192399;			
	RA	Scott D.A., Wang R., Kreman T.M., Sheffield V.C., Karnishki L.P.;			
	RT	"The Pendred syndrome gene encodes a chloride-iodide transport			
	RT	protein.";			
	RL	Nat. Genet. 21:440-443(1999).			
	RN	[4]			
	RP	VARIANTS PDS PRO-236 AND PRO-416.			
	RX	MEDLINE=98282235; PubMed=9618166;			
	RA	van Halbe H., Everett L.A., Coucke P., Scott D.A., Kraft M.L.,			
	RA	Ris-Stalpers C., Bolder C., Otten B., de Vilder J.J.M.,			
	RA	Dierich N.L., Ramesh A., Srisailapathy S.C.R., Parving A.,			
	RA	Cremers C.W.R.J., Willems P.J., Smith R.J.H., Green E.D., van Camp G.;			
	RT	"Two frequent missense mutations in Pendred syndrome.";			
	RL	Hum. Mol. Genet. 7:1099-1104(1998).			
	RN	[5]			
	RP	VARIANTS PDS PRO-236; GLY-384 AND PRO-416.			
	RX	MEDLINE=9828236; PubMed=9618167;			
	RA	Coffey B., Reardon W., Herbrick J.A., Tsui L.-C., Gausden E., Lee J.,			
	RA	Coyne R., Grellet A., Grossman A., Phelps P.D., Luxon L.,			
	RA	Kendall-Taylor P., Scherer S.W., Trembath R.C.;			
	RT	"Molecular analysis of the PDS gene in Pendred syndrome (sensorineural			
	RT	hearing loss and goitre).";			
	RL	Hum. Mol. Genet. 7:1105-1112(1998).			
	RN	[6]			
	RP	VARIANT DEFBA SER-497.			
	RX	MEDLINE=98160177; PubMed=9500541;			
	RA	Li X.C., Everett L.A., Lalwani A.K., Desmukh D., Friedman T.B.,			





Db 613 NAFPERDEIDLEELDLPYKEIETQVOMNSBLPKVNVKPYRPHSLVLDGCAISFLDVG 672

Qy 642 IKALAKLSSTYTGKIGKVFVLNIHAQYYNIDSHGCVFEDSGLECKHYFPSIHDAVLEAQ 700

673 VRSLRVLYKEQRIDVANNVFASLDQDYIEKLEGGCFDD--NIRKDTFFFLVHDAILYLQ 730

```

RESULT 2
DRA_HUMAN
ID DRA_HUMAN STANDARD; PRI; 764 AA.
AC P40879;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DRA PROTEIN (DOWN-REGULATED IN ADENOMA).
GN DRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RX MEDLINE=93248250; PubMed=7683425;
RA Schweinfest C.W., Henderson K.W., Sustar S., Kondoh N., Papas T.S.;
RT "Identification of a colon mucosa gene that is down-regulated in
  colon adenomas and adenocarcinomas.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4166-4170(1993).
RN [2]
RP SIMILARITY TO SULFATE PERMEASES.
RX MEDLINE=94188926; PubMed=8140616;
RA Sandal N.N., Marcker K.A.;
RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
  permease II and a putative human tumour suppressor.";
RL Trends Biochem. Sci. 19:19-19(1994).
RN [3]
RP VARIANTS C1D LEU-124 AND VAL-317 DEL. AND VARIANT TRP-307.
RX MEDLINE=97051927; PubMed=8896562;
RA Hoeglund P., Halla S., Socha J., Tomaszewski L., Saarialho-Kere U.,
RA Karjalainen-Lindberg M.-L., Altiola K., Holmberg C.,
RA de la Chapelle A., Kere J.;
RT "Mutations of the Down-regulated in adenoma (DRA) gene cause
  congenital chloride diarrhoea.";
RL Nat. Genet. 14:316-319(1996).
RN [4]
RP VARIANTS C1D SER-120; ARG-131; VAL-317 DEL AND TYR-527 DEL.
RX MEDLINE=98213471; PubMed=9554749;
RA Hoeglund P., Halla S., Gustavson K.-H., Taipale M., Hannula K.,
RA Poplonska K., Holmberg C., Socha J., de la Chapelle A., Kere J.;
RT "Clustering of private mutations in the congenital chloride
  diarrhoea/down-regulated in adenoma gene.";
RL Hum. Mutat. 11:321-327(1996).
RN [5]
RP FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS SIGNIFICANTLY DECREASED IN
CC ADENOMAS (POLYPS) AND ADENOCARCINOMAS OF THE COLON.
CC -1- DISEASE: DEFECTS IN DRA ARE THE CAUSE OF CONGENITAL CHLORIDE
CC DIARRHEA (C1D), A DISEASE CHARACTERIZED BY VOLUMINOUS WATERY
CC STOOLS CONTAINING AN EXCESS OF CHLORIDE. THE CHILDREN WITH THIS
CC DISEASE ARE OFTEN PREMATURE.
CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L02785; AAA58443.1; -.
DR MIM, 126550; -.
DR MIM, 214700; -.

```

DR	InterPro:	IPR002645;	S7AS.
DR	InterPro:	IPR001903;	Sulfate_transp.
DR	Pfam:	PF01740;	S7AS; 1.
DR	Pfam:	PF00916;	Sulfate_transp. 1.
DR	PROSITE,	PS01130;	SULFATE_TRANSP. 1.
KM	Transports;	Transmembrane;	Disease mutation; Polymorphism.
FT	TRANSMEM	77	97
FT	TRANSMEM	100	120
FT	TRANSMEM	125	145
FT	TRANSMEM	176	196
FT	TRANSMEM	198	218
FT	TRANSMEM	259	279
FT	TRANSMEM	286	306
FT	TRANSMEM	343	363
FT	TRANSMEM	375	395
FT	TRANSMEM	412	432
FT	TRANSMEM	439	459
FT	TRANSMEM	470	490
FT	TRANSMEM	644	664
FT	TRANSMEM	702	722
FT	VARIANT	120	120
FT	VARIANT	124	124
FT	VARIANT	131	131
FT	VARIANT	307	307
FT	VARIANT	317	317
FT	VARIANT	527	527
SO	SEQUENCE	764 AA;	84504 MW; 654CSCBC2D4121F6D CMC64; /FTID=VAR_007430. /FTID=VAR_007431. MISSING (IN CLD). /FTID=VAR_007432. MISSING (IN CLD). /FTID=VAR_007433. /FTID=VAR_007430. C -> W (PROBABE POLYMORPHISM) /FTID=VAR_007431. MISSING (IN CLD). /FTID=VAR_007432. MISSING (IN CLD). /FTID=VAR_007433. /FTID=VAR_007430.

	Query Match	28.9%	Score 1120:	DB 1:	Length 764:
	Best Local Similarity	34.0%	Pred. No. 2,4e-67:		
	Matches 250:	Conservative 142:	Mismatches 275:	Indels 68:	Gaps
QY	7	RYVVDRAAYSLTTFDDEFKKDRTYRVG-EKLNAPFCSSAKIKAVVFGLLPVLSPMKY	65		
Db	8	QYIARPRVYSTNAEEHMKGRYHKKHKKTFDLHLKVCSSQOKAKRIYLSFPIASWLPAY	67		
QY	66	KIKRYIIPDLIGSGSGSIOYPOGMAFALLANIPAVNGILSPFPLLTYFEELGVHOMVP	1255		
Db	68	RLKEWLLSDIYSGISGTIVAVLQSLARALLVDIPVYGLIASFPFALITLFFPGSRISIV	1277		
QY	126	GTFPAVSIIVG---NICQLAPESKPOFV---NNATNESYVDTPAMEAERLHVASLTAC	1787		
Db	128	GPFFILSMWVGLAVSAGVASKAVPDRNATTLGLPNNSSNSSLDD--ERVAVAAASVTV	1848		
QY	179	LTAIIOMKLGCMQGFVAIYISESFINGKMTAGLOLILSVLKIIEGLTITPSTYGPSSIV	2388		
Db	185	LSGIIOIAPFLGRIGFVYISELSIGFTTAAVAVHVLVSOLKIFOLTVYSHTPSIF	2444		
QY	239	FTFIDICKNPHPTNASTLFIALLISGAFVLVAKELNARIMHIREPTEMIVVAVTAIS	2988		
Db	245	KVLXSVFSQIKETINADLVTLALLIYLVVLSYIKEINQRFKDLPRPIIEFIEMVIAAGVS	3048		
QY	299	GGCKMKPKRYHMOIVGEIORGFPPTVSPVSPQMKDMIGTAFSLAIYSYVINLAMGRTLANK	3587		
Db	305	YGCDFKKNFKAAVAVGDMMPGQFPPTTDFVEFQVMTGVCDFGIADVAVAFVSAVSYSLK	3646		
QY	359	HGYVDVDSNOEMIALIGCSNFFSSEFKIHIVICALSVTLVAODSAGGSQVASCYSLVAMIT	4188		
Db	365	YDYPDLGDMOELIALIGIIVCGVIRGFAGSTALRSASVOESTGKTQIAGLIGAIIVLV	4228		
QY	419	MLVIGIYLYPLPKSVLGAFLIIVNLKNSLKQTLDDYVLYMRKSKLDCCTWVVSFLSSPFLSL	4787		
Db	425	VLAIGFLIAPLOKSVLALALGNLKMGMQFAEIGRLMRKQKYDCLLIMIMEFTIYVLGL	4848		
QY	479	PYGAAGVAESVLVVPOTQFRNGTALAAQVMDTIIYVNPKTYNRAODIGIKITITYCSPL	5388		

Db 485 GGLASAVAFOLLITVEKQFPCSTLANIGRTNKKKDYDYMEPEGKIFPCSPFI 544  
 QY 539 YFANSEIFROKVI-----AKTVSLOELQ-----DFENAP 568  
 Db 545 YFANIGFPRKILDAVGFSPRLRKRLKRLKRLKQGLQVTPKFCIGVDIKXS 604  
 QY 569 PFDPPNNQ-----TPANGSVSV-ITFSPDSSSPAOSEPPASAEAGDEPSDMLASVPPV 622  
 Db 605 DEELDNNOIEVLDOPIINTDLEPHIDMNDL-----PLNIEVPR-----I 644  
 QY 623 TFEHTLLDMSGVFFVDMGLKALAKLSYTGKIGVFLVNIHAQVYNDISHGVFEDGS 682  
 Db 645 SLHSLILDSAVSFLDVSSVRLKSLLOEFLIKVDYITGTDDEFTEKLNREEF-DGE 703  
 QY 683 LECKHVPFSDHDAVL 697  
 Db 704 VKSIEFFLTHDAVL 718

RESULT 3  
 DTG\_RAT STDARDO: PRT: 739 AA.  
 ID DTG\_RAT  
 AC 070531:  
 DT 15-JUL-1999 (Rel. 38, Last Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SUIPATE TRANSPORTER (DIASTROPHIC DYSPLASIA PROTEIN HOMOLOG).  
 GN SLC26A2 OR DTDST.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Osteosarcoma;  
 RX MEDLINE=98241598; PubMed=9575183;  
 RA Satoh H., Suuaki M., Shukunani C., Iyama K., Negoro T., Hiraki Y.;  
 RT "Functional analysis of diastrophic dysplasia sulfate transporter.  
 RT Its involvement in growth regulation of chondrocytes mediated by  
 RT sulfated proteoglycans.";  
 RL J. Biol. Chem. 273:12307-12315(1998).  
 CC - FUNCTION: SUIPATE TRANSPORTER. MAY PLAY A ROLE IN ENDOCHONDRAL  
 CC BONE FORMATION.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC - TISSUE SPECIFICITY: CARTILAGE AND INTESTINE.  
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D82883; BA25987.1; -  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate.transp.  
 DR Pfam: PF01740; STAS; 1.  
 DR Pfam: PF00916; Sulfate.transp. 1.  
 DR PROSITE: PS01130; SULFATE\_TRANS. 1.  
 KW Transport; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 111  
 FT TRANSMEM 112 132 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 133 136 POTENTIAL.  
 FT TRANSMEM 137 157 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 158 159 POTENTIAL.  
 FT TRANSMEM 160 180 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 181 218 POTENTIAL.  
 FT TRANSMEM 219 239 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 240 241 POTENTIAL.  
 FT TRANSMEM 242 262 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 263 296 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 297 317 POTENTIAL.  
 FT DOMAIN 318 328 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 329 349 POTENTIAL.  
 FT DOMAIN 350 377 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 378 398 POTENTIAL.  
 FT DOMAIN 399 420 POTENTIAL.  
 FT TRANSMEM 421 441 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 442 454 POTENTIAL.  
 FT TRANSMEM 455 475 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 476 523 POTENTIAL.  
 FT DOMAIN 524 544 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 545 544 POTENTIAL.  
 FT DOMAIN 644 664 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 665 739 POTENTIAL.  
 FT CARBOHYD 205 205 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 357 357 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SQ SEQUENCE 739 AA: 82027 MW: 0C539F66D478C8DA CRC64.

Query Match 25.2%; Score 974.5; DB 1; Length 739;  
 Best Local Similarity 31.7%; Pred. No. 1.2e-57;  
 Matches 219; Conservative 140; Mismatches 280; Indels 51; Gaps 8;

QY 33 VBEKLRNAPRGSASIKAVFGLPLVSLPKYTKINDYIIPDLGLSGSGSIQVPOGMAF 92  
 Db 70 VNRKLOKSCQCNATKIRNRIFFDFPVLRLPKYDKKNILGDMGSLIVGILLVPOSIAY 129  
 QY 93 ALLANIPAVNGLYSFFPLTYFFLFGVHQWPGTFEAVSILVGN-----CLOL-A 143  
 Db 130 SILAOGEPYIGLYTFEFASIIYFLGTSNHSIVGIFGLICLMIGEVDRELHKAPODIT 189  
 QY 144 PESKFOVENN--ATNESYVDTAAMEABERLHVSATLACTALITIOGLFMQGFVAIYLS 200  
 Db 190 TSSSIAMFSGNCVVNHTLDGLCDKSCYAIIKIGSYTFEMAGVQYAMGFQGVSVYLS 249  
 QY 201 EEFINGFMNAGLOLLISLKYIEGLTIPSYGPRSIYFTFDCKNLPHTNIALIFAL 260  
 Db 250 DALISGFTVGASFTILTQAKYLLISLPRNSGVSVITWTHIRNKHKNICOLLTSL 309  
 QY 261 ISGAFVLVLELNARYMHKIRPIPTENIVVVAIAISGCGMPKRYHQVIGEIORGFP 320  
 Db 310 LCLLVLPYRKELNEFKSKLAPITELIVVAATLASHFGKLNINYSIAGQIPTGFM 369  
 QY 321 TPVSPVVSQWKMDIGTAFSLAIVSYVINLAMGRITLANKHGVDVDSNOEMIALGCSNFGS 380  
 Db 370 PQAPDWSLIPVAVDAIAISIIIGFAITVSLSEMPAKKHGTYVKANQEMVAIGFCNIPS 429  
 QY 381 PFKIHVICALSVTLANVGAGKSOVASLCSVYIMTMVIGITLYPLKPSVIGALLAY 440  
 Db 430 FPHCITTSNAALAKTLVKSTGQOTLSAIVSLVLLVLLAPLFLYSLOKCVLGVITIV 489  
 QY 441 NLKNSLKOLTDPYVLMRKSKLDCCIMVVSFLSPFLSLPYGAVGVAFSVLVVPOFOR 500  
 Db 490 NLRGALLFRDLPKMKRSLRMDYIMFVMTLSALLSTREIGLIVGVCSCMVCVIARTOMP 549  
 QY 501 NGYALQVMDTDIYVNPRTYNRAODIGIKITTYCSPLYFANSEIFROKVIKATVSLQEL 560  
 Db 550 KISLIGLEESSEIFESISTYKMLRSKSGIKVRFIAPLYYIKKFCALYKKTLPNPLV 609  
 QY 561 QODEFNAPTDPPNNQNTANGSVSYITFSPDSSSPAOSEPPASAEAGDEPSDMLASVPP 620  
 Db 610 KAAMKKAARKRLKEET-----VTFHGD-----PDEVSMQLSHDP- 643  
 QY 621 FVTFTLLDMSGVFFVDMGLKALAKLSYTGKIGVFLVNIHAQVYNDISHGVFED 680  
 Db 644 -LELHTVVIDCSAIOFLDTAGIHTLKEVRRDYEAIGIOVLLAQCPNPSVDSLANGEY---- 699  
 QY 681 GSLECKH-----VFPFSDHDAVLAFAQANARD 705  
 Db 700 ----CKKREENILFPYSLEAFAFAEESQKE 725

RESULT 4

DTD_HUMAN	STANDARD;	PT;	739 AA.
AC	PS0443;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	SULFATE TRANSPORTER (DIASTROPHIC DYSPLASIA PROTEIN).		
GN	SIC26A2 OR DTDS1 OR DTD.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95007757; PubMed=7923357;		
RX	Haestbacka J., de la Chapelle A., Mahtani M.M., Clines G.,		
RA	Reeve-Daly M.P., Daly M., Hamilton B.A., Kusumi K., Trivedi B.,		
RA	Maeyer A., Coloma A., Lovett M., Buckler A., Kaitila I., Lander E.S.;		
RT	"The diastrophic dysplasia gene encodes a novel sulfate transporter:		
RL	positional cloning by fine-structure linkage disequilibrium		
RL	mapping.";		
RL	Cell 78:1073-1087(1994).		
RN	[2]		
RP	VARIANTS ACG-1B VAL-340 DEL; ASP-425 AND VAL-678.		
RX	MEDLINE=96122050; PubMed=8528239.		
RA	Superti-Furga A., Haestbacka J., Wilcox W.R., Cohn D.H.,		
RA	van der Harten H.J., Rossi A., Blau N., Rimoin D.L., Steinmann B.,		
RA	Lander E.S., Glitzelmann R.;		
RT	"Achondrogenesis type IB is caused by mutations in the diastrophic		
RT	dysplasia sulphate transporter gene.";		
RL	Nat. Genet. 12:100-102(1996).		
RP	[3]		
RP	VARIANTS AO-II GLU-255; TRP-279 AND VAL-715.		
RX	MEDLINE=96152121; PubMed=8571951;		
RA	Haestbacka J., Superti-Furga A., Wilcox W.R., Rimoin D.L., Cohn D.H.,		
RA	Lander E.S.;		
RT	"Aletostegenesis type II is caused by mutations in the diastrophic		
RT	dysplasia sulphate-transporter gene (DTDS1): evidence for a phenotypic		
RT	series involving three chondrodysplasias.";		
RL	Am. J. Hum. Genet. 58:255-262(1996).		
CC	-1- FUNCTION: SULFATE TRANSPORTER. MAY PLAY A ROLE IN ENDOCHONDRAL		
CC	BONE FORMATION.		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
CC	-1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.		
CC	-1- DISEASE: DEFECTS IN DTDS1 ARE THE CAUSE OF DIASTROPHIC DYSPLASIA		
CC	(DTD), AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY		
CC	OSTEOCHONDRODYSPLASIA WITH CLINICAL FEATURES INCLUDING DWARFISM,		
CC	SPINAL DEFORMATION, AND SPECIFIC JOINT ABNORMALITIES.		
CC	-1- DISEASE: DEFECTS IN DTDS1 ARE THE CAUSE OF ACHONDROGENESIS TYPE IB		
CC	(ACG-1B). ACG-1B IS A RECESSIVELY INHERITED CHONDRODYSPLASIA		
CC	CHARACTERIZED BY EXTREMELY POOR SKELETAL DEVELOPMENT AND PERINATAL		
CC	DEATH.		
CC	-1- DISEASE: DEFECTS IN DTDS1 ARE THE CAUSE OF ALETOSTEGENESIS TYPE		
CC	II (AO-II), ALSO KNOWN AS NEONATAL OSSOUS DYSPLASIA I, WHICH IS		
CC	CHARACTERIZED BY SEVERELY SHORTENED LIMBS, SMALL CHEST, SCOLIOSIS,		
CC	CLUB FOOT OF THE EQUINOVARUS TYPE (PALPES EQUINOVARUS), ABDUCTED		
CC	THUMBS AND GREAT TOES, AND CLEFT PALATE. PATIENTS DIE OF		
CC	RESPIRATORY INSUFFICIENCY SHORTLY AFTER BIRTH BECAUSE OF THE		
CC	COLLAPSE OF THE AIRWAYS AND PULMONARY HYPOPLASIA DUE TO THE SMALL		
CC	RIB CAGE.		
CC	-1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
DR	EMBL: U14528; AAA70081.1; -		
DR	IRM: 222600; -		
DR	IRM: 600972; -		

[illegible]



```

Db 550 KNSLGLSESEFESISTYKNIKRSKGIVKRFIAPLYIYINKECKRSALYKALNPVLV 609
QY 561 QODFENAPPTDNNQTPANGTSVSYTFSPDSSPPAOSPPASAEAPGEPDMLASVPP 620
Db 610 KAAMKKAARKRLKEEM-----VFTRGD-----PDEVSNQLSHDP- 643
QY 621 FYTHFTLLIDMSGVSVVDLMGKALKALSLSTYKIGVAVLVNHAQYNDISHGVFEED 660
Db 644 -LEVHTIVDCSAIOFLDPAGIHTHLEKVRDYEAQVIOVLLAQCNPSVDRSLARGEY--- 699
QY 681 GSLECK-----VFPSIDHVALEFAO 700
Db 700 ----CKKEETLEFTSLSEVAFAP 720

RESULT 6
SAT1_RAT
ID SAT1_RAT STANDARD: PRT: 703 AA.
AC P45380;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER)
DE (SULFATE/CARBONATE ANTI-PORTER).
GN SAT-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RA MEDLINE=94132077; PubMed=8300633;
RX Bissig M., Hagenbuch B., Stieger B., Koller T., Meier P.J.;
*Functional expression cloning of the canalicular sulfate transport
RT system of rat hepatocytes."
RL J. Biol. Chem. 268:3017-3021(1994).
CC - FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE. ACCEPTS OXALATE, BUT
CC NOT SUCCINATE AS A CO-SUBSTRATE.
CC - TISSUE SPECIFICITY: LIVER, KIDNEY. LESS ABUNDANT IN MUSCLE AND
CC BRAIN.
CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L23413; AAA17545.1; -
DR InterPro: IPR002645; STAS.
DR InterPro: IPR001902; Sulfate_transp.
DR Pfam: PF01740; STAS; 1.
DR PROSITE: PS01130; SULFATE_TRANSP; 1.
KM Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 76 96
FT TRANSMEM 98 118
FT TRANSMEM 187 207
FT TRANSMEM 209 229
FT TRANSMEM 260 280
FT TRANSMEM 294 314
FT TRANSMEM 341 361
FT TRANSMEM 381 401
FT TRANSMEM 418 438
FT TRANSMEM 486 506
FT TRANSMEM 606 626
FT CARBOHYD 158 158
FT CARBOHYD 163 163
FT CARBOHYD 387 387
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

SQ SEQUENCE 703 AA: 75447 MW: A90D839D70BD2738 CRC64;
Query Match 22.0%; Score 853; DB 1; Length 703;
Best local similarity 29.3%; Pred. No. 1,4e-49;
Matches 201; Conservative 143; Mismatches 304; Indels 38; Gaps 8;
QY 36 KLRNAPRCSSAKIRAVFGLPLVSLMPKXKIDYIIPDLLGLSGSIQVPGMAFALL 95
Db 34 RLKSCSTCMPCAOALVOGLFPIVRLMPQYRLKEYLAGDVMSGLVIGIILVPOAIAYSL 93
QY 96 ANPVAINGLYSSFFPLTYFFLGVHQMVRGTFAVISILVGNIC---LQLA----- 143
Db 94 AGLOPITYSLTYSEFANLTYFLMGTSRHVNVGIFSLCLMGVQVVDRELQLAGPDSDSL 153
QY 144 -PESKFOVFENN-ATNESYVDVTAMAEARLVSATLACTFATIIOMGLFMOFGFVATYSE 201
Db 154 GPGNNSTLNTNTATLVYGLDQDCGRDCHAIATATLTMADLYQWMLKLGIVSTYLSQ 213
QY 202 SFIRGEYTAGIQLISLVKIFGLTIPSYTPGSGIVTFPIDICKNLPHTNISLIFALI 261
Db 214 PLIDGFAMGASVYILTSQARKHLGVRIPIRHQGLGMVHTMLSLQNVGOANLDDVYTSAV 273
QY 262 SGAFVLVYKRLNRYMKIKFPIPTEMIVVYVATISGCKMKPKKYMQVGEIQRGFT 321
Db 274 CLAVLLTAKELSDRYRYLYKVPVTELLIVAVTIAHFGOLHTRFGSSVAGNIPGEVA 333
QY 322 PVPVVSQWMDMIGTAFSLAIVSVINLAMGRTLANKGYDVDSNOBMAILGCSNFGSF 381
Db 334 POIPDKRIMSVALDAMSLVLSGSAFISLAEMFARSHGTSVANQDELLAVGCCNVLP 393
QY 382 FKIHVICALSVTLAVDAGKQSVASLVSVMITMLVIGLYLPRKSVLGAALAVN 441
Db 394 FHEFATSAALSKTLVYKATCGQFQLSSVSAVAVLLVTLVLAFLPHLDQCYLACTIVS 453
QY 442 LKNSLKQUTPYTLMKRSKIDCCIVWVSFLSPFLSPYGVANGVASVLYVYQYQFRN 501
Db 454 LRQALRKVKDLPQLMWRISPDALVWATATCVAVSTEAQLLGVFSLSLAGRQRP 513
QY 502 GYVLAQVMDIDIVNPPTRYRAODIQIKITCYSPFYANSEIFRQKVIATKYSLOEQ 561
Db 514 AALLARIQDSTFYEDAAEFGLLPPEVRYFRFGPLTYANKDF-----LRSLXSLGLD 569
QY 562 QDFENAPPTDNNQTPANGTSVSYTFSPDSSPPAOSPPASAEAPGEPDMLASVPP 621
Db 570 AGY-SATRKROGREVGVSNNSLVDRKDLGSVSS-----GDGLV-VPLA 610
QY 622 VTHTFTLLIDMSGVSVVDLMGKALKALSLSTYKIGVAVLVNHAQYNDISHGVFEED 680
Db 611 FGHHTVVIDCAPLFLFDVACMATLKLRLKRYRALDITLLACCPSPVDRTLRKGGFLGED 670
QY 681 GSLECKHVPSPIDHVALEFAQANRDV 706
Db 671 QGTAEELLPPSVHSAVETACAREEL 696

RESULT 7
YLUM CAEEL
ID YLUM CAEEL STANDARD: PRT: 758 AA.
AC Q94225;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE HYPOHETICAL 85.0 KDA PROTEIN F41D9.5 IN CHROMOSOME X.
GN F41D9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Fulton B.;

```

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN-BRISTOL NZ;  
 RA Waterston R.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: POSSIBLE SULFATE TRANSPORTER.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U67954; AAB52608.1; -  
 DR WormPep: F41D9.5; CE10242.  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; STAS: 1.  
 DR Pfam: PF00916; Sulfate\_transp: 1.  
 DR PROSITE: PS01130; SULFATE\_TRANSP: 1.  
 KM Hypothetical protein; Transmembrane; Transport.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 109 129 POTENTIAL.  
 FT TRANSMEM 192 212 POTENTIAL.  
 FT TRANSMEM 228 248 POTENTIAL.  
 FT TRANSMEM 270 290 POTENTIAL.  
 FT TRANSMEM 294 314 POTENTIAL.  
 FT TRANSMEM 335 355 POTENTIAL.  
 FT TRANSMEM 375 395 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 473 493 POTENTIAL.  
 FT SEQUENCE 758 AA; 84958 MW; 721B87729F2E12P3 CRC64;  
 SQ  
 Query Match 15.28; Score 588; DB 1; Length 758;  
 Best Local Similarity 24.68; Pred. No. 7.6e-32;  
 Matches 175; Conservative 141; Mismatches 295; Indels 100; Gaps 18;  
 QY 35 EKLNAFRCSSAKIKAVFGLPLVSLPKYKIKDYIIPDLGLSGSGSIOVPGMAFAL 94  
 DB 22 QKLR-YACSPKCIHLSFLPIITWLPKDKMSHSFFGDLSSGLTMAVPEPGIALAS 79  
 QY 95 LANLPAVNGLYSFFPLITLYFFELGCVHOMVPGTFAVISIL-VGNICQLA----- 143  
 DB 80 ITGVPPVGYLTATFPPFLYFFEGTSKHNALGCPAVLSLMTGALIEKVMLEATSYNATA 139  
 QY 144 -----PESKFOVFNAT-----NE-SYVDTAAME-----AERLIVASATLA 177  
 DB 140 YVNHITLDELLOKEENTLALISTTLMQILGNFTSVEEVTEMMTEGTVPAVQIHVATTII 199  
 QY 178 CLTAIIOMGLCFMOPGFVAIVLSESRFGFTAGLQILISLVLYRIGLTPSTGPGSI 237  
 DB 200 FLAGVIOVPMGVFRLOLYTSLFSQVMSGCVGGINHVFRAQIOMGLIEPRSGGEYL 259  
 QY 238 VETFDICKNLPHNTNIALSLFALISGAFVLVKELNARYMKH-IRPDIPEMIVVVATA 296  
 DB 260 YRIMLDVENLDNVHIFVTCISLSSFLVFGREYLA PWLNSANVVPPELVILT----- 315  
 QY 297 ISGCKAPKRYHMOIYVEIORGFPTPVSPVVSQKDKMIGAFSLAIVSYINLAMGRTLA 356  
 DB 316 -----HYLI-----NRPPPSLPREDLIRHIGLNAAMAIAITRAVALHITVAAYVE 359  
 QY 357 NKHGVDVDSNOEMTALGCSNPFSGFEKIHVICALLSVTLAVDAGGGSQVSLCVSLVVM 416  
 DB 360 KRYTKINHGOELYALGCVGLVSFFPVFPVTSGFARSVCAAVGSGTOLTCLPSSIALL 419

QY 417 ITMLVGLIYLPKPSVGLALIAVNLKSLKQLTDPYLMRKSRLDCCINVVSLSFEFL 476  
 DB 420 SVIICIGPALLEYLPQCLISAMIFIAQKGMLEKFGELSLMNVFPIIDFTIMLSFEFLVCY 479  
 QY 477 SLPIYGVAVAFSVLVVVFQTFRNGVYALAOVMTDIYVNPKTNRADODIG-IKIITYC 535  
 DB 480 DMGGLMALIOFAVLTITIRORRWMFLSRDDETENYKTKRRD-LERIQGNVCIFRMD 538  
 QY 536 SPLVFANSEIRQVAVIAKTVSLQELQDFENAPPTDPNNQTPANGTSVSYIT---FSPD 592  
 DB 539 APLFTSSDR-----TKSVQCVKMKERC-----KSESVTIEQANSND 577  
 QY 593 SSSPAQSEPPASAE---APGEPDMLASVPPVTFHLLIDMSGVSFVDMGLAKLKS 649  
 DB 578 RSADIFDSKLSARRRMRKRDQSENRC-----LVIDCGFPVYDGLSTLSKVY 628  
 QY 650 STYKIGYKVLVNIHNAVYNDISHGVFEQGSLECKHVPFSIHDAVLAFAQ 700  
 DB 629 VDLQAAGIQCFV-VKSDKLKLRATDEYEVDESK-VFNKVGDAVKAKE 677  
 RESULT 8  
 SUT2\_STYHA STANDARD; PRT; 662 AA.  
 AC P53392;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HIGH AFFINITY SULPHATE TRANSPORTER 2.  
 GN S72.  
 OS Stylisanthes hamata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;  
 OC Stylisanthes.  
 OX NCBI\_TaxID=37660;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. VERANO; TISSUE=Root;  
 RX MEDLINE=96016171; PubMed=7568135;  
 RA Smith F.W., Ealing P.M., Hawkesford M.J., Clarkson D.T.;  
 RT "Plant members of a family of sulfate transporters reveal functional  
 RT subtypes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).  
 CC - FUNCTION: HIGH-AFFINITY H+/SULFATE COTRANSPORTER THAT MEDIATES THE  
 CC UPTAKE OF SULFATE BY PLANT ROOTS FROM LOW CONCENTRATIONS OF  
 CC SULFATE IN THE SOIL SOLUTION.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X82256; CA57711.1; -  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; STAS: 1.  
 DR Pfam: PF00916; Sulfate\_transp: 1.  
 DR PROSITE: PS01130; SULFATE\_TRANSP: 1.  
 KM Transmembrane; Transport; Multigene family.  
 FT TRANSMEM 103 123 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 182 202 POTENTIAL.  
 FT TRANSMEM 205 225 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 291 311 POTENTIAL.  
 FT TRANSMEM 346 366 POTENTIAL.

FT TRANSMEM 383 403 POTENTIAL.  
 FT TRANSMEM 420 440 POTENTIAL.  
 FT TRANSMEM 447 467 POTENTIAL.  
 FT TRANSMEM 481 501 POTENTIAL.  
 SO SEQUENCE 662 AA; 72728 MW; 7544DLE73FB76284 CRC64;

Query Match 13.4%; Score 518; DB 1; Length 662;  
 Best Local Similarity 26.0%; Pred. No. 3e-27;  
 Matches 183; Conservative 121; Mismatches 277; Indels 122; Gaps 22;

18 TLPDDEFEKKDRY----EYGEKLNARFRCSSAKIKAVFGL---LPVLSMLPKYIKDY 70  
 Db TLFOELKHSNFNEFFPKDQSGFR-----FKDQSSRK-FVLGLQYIPILDMGRHNDLKR 101  
 71 IIPDLGLSGSGSIQVPCGMALLANLPAVNGLYSFFPLTYFFLGVHOMVPGFAY 130  
 Db 102 -KGDPLAGLTILSLCIPDIAAKLANLDPWGLYSSFAVPLVAFMGTSRDIAIGFAY 160  
 131 ISLVGNICLOLAPESKFOVFNATNESYVDTAAMEAEHLVASATLACTAIIONGLGFM 190  
 Db 161 VSLILGLT---LSNE-----ISMTKSHDYL-----RLAFATF--FAGVYQMLLGVC 202  
 191 QGFAVIYIYSESTIRGFMFAGIQLIISVKYIFGLTIPSYGPGSIVTFPIDCNLPH 230  
 Db 203 RLGLFLDPLSHAIYGFMAAGAITIGLOOLKGLG--IKDFTKNSDIVSMHVSVMVHH 260  
 251 -TNIASLIPALISGAFVLVYKELNARYMKIRPIPTEMIVVAVATLAISGCKMPKKYHM 309  
 Db 261 GNMWETILGLSLFLITLTKYIAKNNKLEFWSAISPMICVISTFEVYITRADKR-GV 319  
 310 QIVGELIORGF-PTVSPVVSQKMDMIGTAFSLAIVSYNL---AMGRTLANKHGYDND 364  
 Db 320 TIVKHIKSGVNPSSANEIFFHGR-YLGAGVRGVAVGLVATEMAIGTFEAMKDYSD 378  
 365 SNOEMALICGSNFGSFFRIHVICALSVTLANDAGKSOVASLCSVLYMTLVIGI 424  
 Db 379 GKNEMVAMGTMNIVGSLTSCYVTGFSRSVAVVMAGCKTAVSNIVAIYVLLTLVITP 438  
 425 YLYPLKSVLGLI---AVNLKNSLKOLDPYLYLMRSKIDCCIMWVSFLSFFLSPLYG 461  
 Db 439 LKRYTNNAVLASTIILAAVNVLN-----IEAMVLLMKIDKDFACMGAFEGVIFKSEIG 494  
 482 VAVGAFSVLVVVFQTFRNGVALAQMVDIYVNPXTYNNRAODIGIKLITYCSPLYFA 541  
 Db 495 LILAVIAISPAKILLOYTRPTAVLGKLPSTSVRNIOQYKABQIPOMLIIIRVDSAIYS 554  
 542 NSEIFQKIYAKIVSLOELODFENAPPTDPNNOPTPANGTSVYITFSDDSSPAQSEP 601  
 Db 555 NSWYIKERILMLID-----EGAQRTESL 579  
 602 PASAEPAGEPSDMLASVPPEVTFHTLIDMSGVFVDMGKALKLSTYYGKIGVFL 661  
 Db 580 P-----EIOHLIVEMSPYTDIDTDSIHAFELIKYLQKREYQML 619  
 662 VN-----HAQVYNDISHGVFEDGSLCKHVPESIDAV 696  
 Db 620 ANGPVYIEKLHASNLAEL-----IGED-----KIFLVADAV 652

RESULT 9  
 SUTL\_STYHA STANDARD; PRT; 667 AA.  
 AC P53391;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HIGH AFFINITY SULPHATE TRANSPORTER 1.  
 GN STL.  
 OS Stylosanthes hamata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;

OC Stylosanthes.  
 ON NCBI\_TaxID=37660;  
 RX (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. VERRANO. TISSUE=Root;  
 RX MEDLINE=96016171; PubMed=7568135;  
 RA Smith F.W., Ealing P.M., Hawkesford M.J., Clarkson D.T.;  
 RT "Plant members of a family of sulfate transporters reveal functional  
 subtypes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).  
 CC -! FUNCTION: HIGH-AFFINITY H+/SULFATE CO-TRANSPORTER THAT MEDIATES THE  
 CC UPTAKE OF SULFATE BY PLANT ROOTS FROM LOW CONCENTRATIONS OF  
 CC SULFATE IN THE SOIL SOLUTION.  
 CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -! SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X82555; CAA57710.1; -  
 CC InterPro: IPR002645; STAS.  
 CC InterPro: IPR001902; Sulfate\_transp.  
 CC Pfam: PF00916; Sulfate\_transp. 1.  
 CC PROSITE: PS01130; Sulfate\_transp. 1.  
 CC Transmembrane; Transport; Multiligene family.  
 CC TRANSMEM 106 126 POTENTIAL.  
 CC TRANSMEM 131 151 POTENTIAL.  
 CC TRANSMEM 156 176 POTENTIAL.  
 CC TRANSMEM 185 205 POTENTIAL.  
 CC TRANSMEM 208 228 POTENTIAL.  
 CC TRANSMEM 269 289 POTENTIAL.  
 CC TRANSMEM 296 316 POTENTIAL.  
 CC TRANSMEM 350 370 POTENTIAL.  
 CC TRANSMEM 425 445 POTENTIAL.  
 CC TRANSMEM 452 472 POTENTIAL.  
 CC TRANSMEM 486 506 POTENTIAL.  
 CC SEQUENCE 667 AA; 73172 MW; 1A7AD47FDEB4DA7E CRC64;

Query Match 13.3%; Score 516; DB 1; Length 667;  
 Best Local Similarity 25.5%; Pred. No. 4.1e-27;  
 Matches 179; Conservative 128; Mismatches 278; Indels 116; Gaps 20;

18 TLPDDEFEKKDRY---PVGE-KLNARFRCSSAKIKAVFGLLPVLSMLPKYIKINDYII 72  
 Db TLFOELKHSNFNEFFPKDQSGFR-----KLELGLQYIPILDMGRHNDLKR-R 105  
 73 PDLGLSGSGSIQVPCGMALLANLPAVNGLYSFFPLTYFFLGVHOMVPGFAYIS 132  
 Db 106 GDFPLAGLTILSLCIPDIAAKLANLDPWGLYSSFAVPLVAFMGTSRDIAIGFAYVS 165  
 131 ISLVGNICLOLAPESKFOVFNATNESYVDTAAMEAEHLVASATLACTAIIONGLGFMQF 192  
 Db 166 LILGLT---LSNE-----ISMTKSHDYL-----RLAFATF--FAGVYQMLLGVCRL 207  
 193 GFVAIYIYSESTIRGFMFAGIQLIISVKYIFGLTIPSYGPGSIVTFPIDCNLPH-T 251  
 Db 208 GLFLDPLSHAIYGFMAAGAITIGLOOLKGLGSGSNNEFKKEDLISVMRSVMTHVHGW 267  
 252 NIASLIPALISGAFVLVYKELNARYMKIRPIPTEMIVVAVATLAISGCKMPKKYIMO 311  
 Db 268 NMWETILGLSLFLITLTKYIAKNNKLEFWSAISPMICVISTFEVYITRADKR-GVSI 326  
 312 VGEIORGFP-PTVSPVVSQKMDMIGTAFSLAIVSYNL---AMGRTLANKHGYDVDSN 366  
 Db 327 VKHIKSGVNPSSANEIFFHGR-YLGAGVRGVAVGLVATEMAIGTFEAMKDYALDGN 385

QY 367 QEMIALGCSNFEFSFKIHVICALSVTLAVDAGGKSQVASCVSLVMTMLVLGILY 426  
 DB 386 KEMAMGTMTNIVGSLSSCYVTTGFSRSANVYMGAKRAVSNVIMSVILLTLVTLPLF 445  
 QY 427 YPLKSVGLI---AVNLKSLKQLDPPYLYMKRSKIDCCIMVVSFLSSFLSPYVA 483  
 DB 446 KYTNNAVLASTIIIAVNLVN---IEAMVLLMKIDKFDVACGAFEGVIFKSVIEGLL 501  
 QY 484 VGVAFSVLVVVFQTFRNGYALAOVMDTIYVNPRTYRRAODIGIKITTCSPLEYFNS 543  
 DB 502 IAVNISFAKILLQYTRPTAVLGKLPQTSYRNIOQYPKAQIGMLIRDSAIYFNS 561  
 QY 544 EIFQOKYIAKTVSLQELQDFENAPPTDPNNOTPANGTSVYITFSPDSSSPAOSEPPA 603  
 DB 562 NYIKERILRWLD-----EGAQRTESELP- 585  
 QY 604 SAAPGERSDMLASVPFVFTHTLLDMGVSFVDMGKALAKLSSTYGGIKVFLVN 663  
 DB 586 -----EIQHLITKSPVDPIDTSGIHAFELYKTLQREVOLILAN 626  
 QY 664 -----IHAQVYNDISHGCVFEDGSLCKHVPFSIHDAV 696  
 DB 627 PGVPIEKLNHASKLTCL-----IGED-----KIFLYADAV 657

RESULT 10  
 SUT3\_STYHA STANDARD; PRT; 644 AA.  
 AC P53393;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE LOW AFFINITY SULPHATE TRANSPORTER 3.  
 GN ST3.  
 OS Stylosanthes hamata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;  
 OC Stylosanthes.  
 ON NCBI\_TaxID=37660;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CY. VERANO; TISSUE=Root;  
 RX MEDLINE=96016171; Pubmed=7568135;  
 RA Smith P.W., Ealing P.M., Hawesford M.J., Clarkson D.T.;  
 RT "Plant members of a family of sulfate transporters reveal functional  
 subtypes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).  
 CC -!- FUNCTION: LOW-AFFINITY H+/SULFATE COTRANSPORTER WHICH MAY BE  
 INVOLVED IN THE INTERNAL TRANSPORT OF SULFATE BETWEEN CELLULAR OR  
 SUBCELLULAR COMPARTMENTS WITHIN THE PLANT.  
 CC -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X82454; CAA57831.1; -  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; STAS; 1. Sulfate\_transp.  
 DR Pfam: PF00916; Sulfate\_transp; 1.  
 DR PROSITE: PS0130; SULFATE\_TRANSP; 1.  
 KW Transmembrane; Transport; Multigene family.  
 FT TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 76 96 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 156 176 POTENTIAL.  
 FT TRANSMEM 179 199 POTENTIAL.  
 FT TRANSMEM 242 262 POTENTIAL.  
 FT TRANSMEM 268 288 POTENTIAL.  
 FT TRANSMEM 328 348 POTENTIAL.  
 FT TRANSMEM 394 414 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 FT TRANSMEM 455 475 POTENTIAL.  
 FT TRANSMEM 518 538 POTENTIAL.  
 SO SEQUENCE 644 AA; 69785 MW; DA2A0D4FDB5ED280 CRC64;

Query Match 12.88; Score 496; DB 1; Length 644;  
 Best Local Similarity 26.48; Pred. No. 8.5e-26;  
 Matches 180; Conservative 127; Mismatches 260; Indels 116; Gaps 26;

QY 41 FRCSAK--IKAYVE--GLPVLMLPKY---KIKDYIIPPLLGSGSIOVQGNMA 93  
 DB 39 FTSSSKKETRAVSFLASLPILSMIRYSATKRD---DLISGLTLASLISFOSIGYA 94  
 QY 94 LLANLPANGLSSFFPLTYEFLGVHQAHPGTFAVISILVGNICQLAPESKFOYENN 153  
 DB 95 NLAKLDPQGLTYSVIPVYIALMGSSREIALGPAAVSMLLSS---LVPK----- 142  
 QY 154 ATNESYDTAAMEARELHVSATLACLPATIDMGICFMQFGVAIYLSFIRGFTANGL 213  
 DB 143 -----VIDPDAPHNDYRMLVFTVTLFAGIFQTAFCVLRGLVDFLSHAALVGMAGAI 197  
 QY 214 QILSLVKYIRGLIPSTYSGSIVFPFIDICKNLPHNINAS-----LIPALISGAFV 267  
 DB 198 VIGLOLKLGLLT--HETTDVAVALKSVYTSI-HQOITSENWSLNF-VIGCSFLI 253  
 QY 268 LKLELMARYM--HKIREPIP--TEMIVVAVTAI---SGCKMKPKYHMOIVEIQGF 319  
 DB 254 FL--LAARFIORNNKKFWMALPALPLSLVILSTLIVFLSKG---DKHGVNIIKHVOGGL 307  
 QY 320 -PTVSPVYSGMKMIGTAFSLAIVSYINL---AMGRITANKHGVDVDSNCKMALG 374  
 DB 308 NPSSVHKIQLNGPR-VGGAARKIGLISALIALTEAIAVGRSPANKRGYHLDNKEMLAMGC 366  
 QY 375 SNFSGFFKIHVICALSVTLAVDAGGKSQVASCVSLVMTMLVGLYLYPLKPSVL 434  
 DB 367 MNIGSLTSCYVSGFSRTAVNFSAGCKTAVSNIVMAVYLLCLELFTLLTYTPMIL 426  
 QY 435 GALLAVNLKSLKQLDPPYLYMKRSKIDCCIMVVSFLSSFLSLPYGAVGAVSVLVV 494  
 DB 427 ASILSLALP-GLIDIGEAHIMKVKDEPLACLAFGVLFVSIEIGLIALSIPAKIL 485  
 QY 495 FQTFRNGYALAOVMDTIYVNPRTYRRAODIGIKITTCOS-PLYFRANSEIFQOKYIAK 553  
 DB 486 LQAIRPGVEVLGRIPTEAYCDVAQYPAVYTPGILVIRISSGSLCFAMAGFVREIRL-K 544  
 QY 554 TVSIQELQDFENAPPTDPNNOTPANGTSVYITFSPDSSSPAOSEPPASAEAPERSD 613  
 DB 545 WVEDEE-ODNIEEA-----AKG----- 560  
 QY 614 MLASVPFVFTHTLLDMGVSFVDMGKALAKLSSTYGGIKVFLVNIHAQVYNDIS 673  
 DB 561 -----RQALIIIDMTDLTNDTSGIILAEELHKKLSRGVLAVMVNRMEVHKLK 611  
 QY 674 HGVFEDGSLCKHVPFSIHDAV 696  
 DB 612 VAN-FVD-KIGKERVFLTVAEAV 632  
 RESULT 11  
 SUT3\_SCHPO STANDARD; PRT; 877 AA.  
 AC 074377;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE SULFATE PERMEASE C3H7.02.



```

GN SPBC3H7.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Jimenez Martinez J.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -! SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL031261; CAA20298.1; -.
CC InterPro: IPR002645; STAS.
CC InterPro: IPR001902; Sulfate_transp.
CC Pfam: PF01740; STAS; 1.
CC Prosite: PS01130; Sulfate_transp; 1.
KM Transpote; Transmembrane.
FT TRANSMEM 133
FT TRANSMEM 161
FT TRANSMEM 181
FT TRANSMEM 186
FT TRANSMEM 206
FT TRANSMEM 221
FT TRANSMEM 241
FT TRANSMEM 243
FT TRANSMEM 263
FT TRANSMEM 292
FT TRANSMEM 312
FT TRANSMEM 329
FT TRANSMEM 349
FT TRANSMEM 384
FT TRANSMEM 404
FT TRANSMEM 424
FT TRANSMEM 444
FT TRANSMEM 461
FT TRANSMEM 481
FT TRANSMEM 484
FT TRANSMEM 504
FT TRANSMEM 518
FT TRANSMEM 538
FT TRANSMEM 543
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 12.0%; Score 464.5; DB 1; Length 877;
Best Local Similarity 24.4%; Pred. No. 1.6e-23;
Matches 197; Conservative 128; Mismatches 339; Indels 143; Gaps 32;

```

```

DB 319 LKRYTKEQRFVFLTNVLRSAVIIIVGTALSYGVCKHRRNPISILGTVPSGRDMGVP 378
OY 326 VWSQ--WKDMIGTAFSAIIVSYVNLAMGRTLAKHGDVDSNOEMTALCCSNFFGSFFK 383
DB 379 VISRKLCAIDLASELPVSVIYVLLLEHISIAKSPGVNRYKVIIPDELIMKATNLIGVFH 438
OY 384 IHWICCALSVTLAVDAGKGSQVASICVSLVMTMLVLCIYLPDKSVGALIAVNLK 443
DB 439 AYPATGFSRSATNAKSGVTRPLGIFTAGVYVALCLTGAFFYIPNAVLSAVIIHSVF 498
OY 444 NSLKQLTDPYLMKSKLDDCICWVVSFLSPFLSPYGVANGVAFSVLVVVFQFQFNG- 502
DB 499 DLITPMRQTLTFMWMQPLEALIFCAVAVSFSSIEINGIYTAVALCLALFLTRKPSGS 558
OY 503 ----YALQVMD-----TDIYV--NPKTYRAADIOGIRIKIITYCSPLYEANSSEIFR 547
DB 559 FLGLIKTANNFDDENSIDVRODIYVPLNKGMM-----PLTYTRDPP--AGVLIIFR 608
OY 548 QKVIATKTVSLQELQODFENAPPTDPNNNOPTPANGTS-VSYIT-----RSPDSSSPAQS 599
DB 609 -----IQESF-----TYPNAGHNSMLTSKAKTVTRGNANIIYKKASDRPMND 651
OY 600 EPASAEAPGEPSPMLASVPPEVTFHTLLIDMSGVSPVDMGKALA---KLSTYKIG 656
DB 652 PAPRKKTNABEVED---TRPLL--RAIILDFSAVNHIDTGTVALVDTRKELEYADE 705
OY 657 VKVELVNIHAQVYND-----ISHG-GVFEDGSLCKHVPFSIDAVLFAQANARDV--- 706
DB 706 VEFHFTDIN---NDWIKRTLVAANGFGKARDAT---KYTSISIE---VGSANPLRDIETP 755
OY 707 -TTEHNFQAGPDAEELSLYSEEDIRS 732
DB 756 MAPGNSRIWPSVVRVRFDEEAIES 782

RESULT 12
ID SULX_YEAST STANDARD; PRT; 754 AA.
AC P53394;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE SULFATE TRANSPORTER YPR003C.
GN YPR003C OR YP9723.03C OR LP23C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Badcock K., Bowman S., Churcher C.M., Pearson D., Rajandream M.A.,
RA Walsh S.V., Barrell B.G.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -! SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 271255; CAA95043.1; -.
CC EMBL: U31900; AAA97582.1; -.
CC EMBL: Z48951; CAA88781.1; -.
CC SGD: S0006207; YPR003C.
CC InterPro: IPR002645; STAS.
CC InterPro: IPR001902; Sulfate_transp.
CC Pfam: PF01740; STAS; 1.

```

DR Pfam: PF00916; Sulfate\_transp: 1.  
 DR PROSITE: PS01130; SULFATE\_TRANS: FALSE\_NEG.  
 KW Hypothetical protein; Transmembrane; Transport.  
 FT TRANSMEM 119 139 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 200 220 POTENTIAL.  
 FT TRANSMEM 233 253 POTENTIAL.  
 FT TRANSMEM 283 303 POTENTIAL.  
 FT TRANSMEM 318 338 POTENTIAL.  
 FT TRANSMEM 371 391 POTENTIAL.  
 FT TRANSMEM 411 431 POTENTIAL.  
 FT TRANSMEM 451 471 POTENTIAL.  
 FT TRANSMEM 475 495 POTENTIAL.  
 FT TRANSMEM 518 538 POTENTIAL.  
 SQ SEQUENCE 754 AA; 83383 MW; 4558DAB879803EEG CRC64;

Query Match 11.38; Score 437; DB 1; Length 754;  
 Best Local Similarity 22.5%; Pred. No. 9e-22;  
 Matches 162; Conservative 138; Mismatches 265; Indels 154; Gaps 23;

QY 56 LPVLWMLPKKIKYIIPDLGLGSGSIQVPGMAFAL-LANPANGLYSSFFLLTY 114  
 DB 105 LPEFSWLPETFE-NKLMGDVIAGISVASFQIPALSTYTAHPPLCGLSLAISPEVY 163  
 QY 115 FFLGGVQWVPGFEPAVSIIVGNICQLAPESKFOVFNMTNESYVDTAAMEARLHVA 174  
 DB 164 GILGSVQWVPGFEPAVSIIVGNICQLAPESKFOVFNMTNESYVDTAAMEARLHVA 174  
 QY 175 TLACIAIIONMGLEWQFGEVAIYLSFIRGEMTAAGLOLISLV-----KYIFGLT 227  
 DB 206 VIFVFSSTILLFSGISFRGELGNLSKALLRGFISVGLVMTIINLSLEKLDKFLVSLP 265  
 QY 228 IPSTGSGIVFPIIDICKNLPHTNINSLFALISGAFVLYVKELNRYM--HKIRFPI 285  
 DB 266 -QHHTPEKILFLIDAPAOYH--IPTALFSGCCLIVLEFLTKRLKLMKRYHSAIFP 322  
 QY 286 TEMIVVAVATAISGCKMPKRYHMOIVGEIQR-GEPTVPVSVVQWMDIGTAS----L 340  
 DB 323 DILLVIVITLLISMKFKLHRYGSIIGDPSMDNFDLKNLPTRPKRLIPDLFSASLIY 382  
 QY 341 AIVSYVNTLAKGRTLANKHGVDNSQDMITALGCSNFFGSF-----KIHVIC 388  
 DB 383 AMGFESTTASKSLGTTYNLTYSNRELVALGFMNTIVISLFGALPAFGYGRSKINAL- 441  
 QY 389 CALSVTLAVDAGGKSOVASLGVMTIMLVIGIYLPKSVLALALAVNLKNSLKQ 448  
 DB 442 -----SGAQSVMSGVEMVITLITMNLLOFVHTIPNCVLSVITITIGISLDEE 490  
 QY 449 LTPDYVLMRSKLDLC-----CIWVVSFLSFLSLPFGVAVGVAVSVLVVVFOTQFRNG 502  
 DB 491 VPGDI-----KFHLRCGGFSELPFAVTFCTTIFYSIAGICIGCVYSIINLIKASARI 546  
 QY 503 YALAQVMDTIYVNPRTY-----NRAODIGIKITTYSPLYFANSEIFRQKVI 551  
 DB 547 QILARVAGTGNFNLDDYMMNKRNSLDVETIEICMIVIRPEPLTFNSELKQRL- 605  
 QY 552 AKTVSIQELQDFENAPPTDPNNQTPANGTSVYITFSDSSSPAQSEPPASAEAGEP 611  
 DB 606 -----DRIERYGSSKIH-----PGRK 621  
 QY 612 S-----DMLASVPEPVYFHTLLDMGSGVSPVDMGKALAKLSITYGKIGKVLVNIHAQ 667  
 DB 622 SLNSKDSIKIVY-----IFDLGWTSIDSSAAQVLEELIITSYKRRNVFIYLVNS- 670  
 QY 668 VYNDISHGCVFEDGSLCKKIVFPISINDAVLFAQANARDVTPGHNFQCAPGDDELSTLYDS 726  
 DB 671 -INDKVARRLFKAG-----VAASVERAQANNNNMTSNTFSDA-GETSPYSDS 717

ID SUL2\_YEAST STANDARD; PRT; 893 AA.  
 AC Q12325;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE SULFATE PERMEASE 2 (HIGH-AFFINITY SULFATE TRANSPORTER 2).  
 GN SUL2 OR SEL2 OR YLR092W OR U9449.1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
 RA Tatch A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
 RA Watson R., Waterston R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Benes V., Rechmann S., Nentwich U., Schwager C., Ansoerge W., Voss H.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97207837; PubMed=9055073;  
 RA Cherest H., Davidian J.C., Thomas D., Benes V., Ansoerge W.,  
 RA Surdin-Kerjan Y.;  
 RT "Molecular characterization of two high affinity sulfate transporters  
 in *Saccharomyces cerevisiae*.";  
 RL Genetics 145:627-635(1997).  
 CC - FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC - SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC -  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U53880; AAB67596.1; -  
 DR EMBL: Z73264; CA97653.1; -  
 DR EMBL: Z73265; CA97653.1; -  
 DR EMBL: U53876; AAB67550.1; -  
 DR SGD: S0004082; SUL2.  
 DR InterPro: IPR002645; STAS.  
 DR InterPro: IPR001902; Sulfate\_transp.  
 DR Pfam: PF01740; STAS: 1.  
 DR Pfam: PF00916; Sulfate\_transp: 1.  
 DR PROSITE: PS01130; SULFATE\_TRANS: 1.  
 KW Transport; Transmembrane; Glycoprotein.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 306 326 POTENTIAL.  
 FT TRANSMEM 351 371 POTENTIAL.  
 FT TRANSMEM 400 420 POTENTIAL.  
 FT TRANSMEM 444 464 POTENTIAL.  
 FT TRANSMEM 484 504 POTENTIAL.  
 FT TRANSMEM 539 559 POTENTIAL.  
 FT TRANSMEM 560 580 POTENTIAL.  
 SQ SEQUENCE 893 AA; 99650 MW; 67826955ACTC0BF5 CRC64;

Query Match 10.5%; Score 405; DB 1; Length 893;



```

QY 117 LGGHOMVPGTFAVISILVGNI---CLQAPESKFOVENNATNESYVDATAMENRLHVS 173
DB 167 FATSQDVCICPVAVMSLQFAKVAIEVLKRYPEDETEV-----TAPI-----IA 209
QY 174 ATACLTATIQMLGFMQGFVAIYSESTIRGFMPTAGIOLISVLKTYFGLTIPSYTG 233
DB 210 TTTCLLCGIVATGTLGIRLFLVELLISLNAVAGFMGSAFNI-----IWG-QIPALMG 261
QY 234 PGSIFF-----FIDICNLPHTNINSLIFALI-----SGAFVLVKEELNAR 275
DB 262 YNLVMTREATYKVVINTLKLHPNTKL-DAVFGILPIVLIVKMKMGCTGITLADYYR 320
QY 276 YMKIRPEPIPT-----EMIVVAVATAISGCC--KMKRYHMOIVGEIORG-PTP 322
DB 321 NQPKRANRLKSFYFAQAMRNNAVIVYFETASITRNKSKDRPISILCTPSGLNEVG 380
QY 323 VSPVVSQMKMIGTAPSLAIVSYI-NLAMGRILANKHGDVDSNOGMIALGCSNFGSF 381
DB 381 VMKIPDGLSNMSEIIPASTIVLLEHIALSKSFGRIKNDKVPDOLLAIGVYNLIGTF 440
QY 382 FKIHVICALSVTLAVDAGAGKSOVASLGVSMITMLVIGIYVLPKSVLCALIAVN 441
DB 441 FHSYPRATGSSRSBALAKKACVNRPFESCVFTGGCVLLALYCLTDAFFIIPKATLSAVIHA 500
QY 442 LKNSLAKQLDPYILMRKSKLDCCITWVSLSPFLSLPYAVAGVAVSVLVVVFQTFRN 501
DB 501 VSDLLTSYKTTWTFMKTNPLDCISFIYVTFVTFSSIEGNYFAMCWSCAMLLKQAFPA 560
QY 502 GVALAQVMDIIVNPTRYRADDIQIKITTCSPILYANSEIFRO-KVIKTVSLQE 559
DB 561 GKTLGRVEVAEV-LNPV---QEDIDAV-----ISSNEIPNLKOVSTVEVLP 606
QY 560 LQO-----DFENAPRTDPNNQOT---PAN-----GTSVSYIFFS----- 590
DB 607 APYKFSVKVAVRPDHCGRSLNINTYRPPRGYIVYRLDOSTFYVCSNHYDIIPRIK 666
QY 591 -----PDS-----SSPAQSEPPAS-AEAPGEP 612
DB 667 EETRRQOLITLRKKSRRPMPNDPEMKKPDLSLKFKRKRSATTPNSDPLTSSNGSETY 726
QY 613 DMLASVPEVTFPHLLIDMSGVSPFDLMGIKALAKLSSTYGKIG-----V 657
DB 727 E-----KPLL--KVCLDFEQVAQVDSAVOSLVLDLRKANRVRADROVEERHAGIISPMI 779
QY 658 KVELVNIHAQVYNDISHGVFEDGSL---ECKHVPESINDAVLFAQANADVTPGHNFQ 713
DB 780 KRSLLSVYKCTGTNE-----EYSDDSIIAGHSSPHAVAKVLADVDYDEDSRISYSNTE 834
QY 714 G--APGDAELSLYDEEDIRSYDWL 736
DB 835 TLCAATGTNLFPFHIDIPDFSKMDV 859

RESULT 15
NO70_SOYBN
ID NO70_SOYBN STANDARD; PRT; 485 AA.
AC 002920;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 01-NOV-1995 (Rel. 32, last annotation update)
DE EARLY NODULIN 70.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. AKISENGOKU;
RX MEDLINE=93241143; PubMed=7683079;
RA Kouchi H., Hata S.;
RT "Isolation and characterization of novel nodulin cDNAs representing
genes expressed at early stages of soybean nodule development.";
```

```

RL MOL. Gen. Genet. 238:106-119(1993).
RN [2]
RP SIMILARITY TO SULFATE PERMEASES.
RX MEDLINE=94188926; Pubmed=8140616;
RA Sandal N.N., Marcker K.A.;
RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
permease II and a putative human tumour suppressor.";
RL Trends Biochem. Sci. 19:19-19(1994).
CC -1- FUNCTION: POSSIBLE SULFATE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT EARLY STAGES OF NODULE
DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: D13505; BAA02723.1; -.
DR PIR: S34800; S34800.
DR InterPro: IPR001902; Sulfate_transp.
DR Pfam: PF00916; Sulfate_transp. 1.
DR PROSITE: PS01130; SULFATE_TRANS. 1.
KW Nodulation; Transmembrane; Transport.
FT TRANSMEM 74
FT TRANSMEM 97
FT TRANSMEM 121
FT TRANSMEM 152
FT TRANSMEM 186
FT TRANSMEM 232
FT TRANSMEM 262
FT TRANSMEM 321
FT TRANSMEM 361
FT TRANSMEM 400
FT TRANSMEM 421
FT TRANSMEM 441
FT TRANSMEM 455
FT TRANSMEM 475
FT TRANSMEM 485
SQ SEQUENCE 485 AA; 52945 MW; 3738B6F64383BCB5 CRC64;

Query Match 9.7%; Score 375.5; DB 1; Length 485;
Best Local Similarity 25.6%; Pred. No. 6,5e-18;
Matches 132; Conservative 98; Mismatches 210; Indels 75; Gaps 13;

QY 1 MSQPR-----RYVDRRAVSGLT-----LEDFEKKDRYPPYGEKLRNFRCSA 46
DB 7 LMAPEPSSMLRQVVDVNYKETLLPHNPNTFSYLNQPSKR-----AF----- 49
QY 47 KIRAVVGLLPVLSMLPKY---KIKDYIIPDLGLSGSGSIQVPOGMAFALLANLPAVNG 103
DB 50 ---ALLONLEPPIIASLQNYNAOKLK---CDIAGLGLALFAIPQCGMNTLARLSPEYG 102
QY 104 LYSSFFPLTYFELGVGHQVPGTFAVISILVGNICLQAPESKFOVENNATNESYVDTA 163
DB 103 LYTGIYVPLIYAMLASREIIVIGPSVDSLILSMIOTL---KPIHDSST----- 150
QY 164 AMEERLHVATLACLTATIQMLGFMQGFVAIYSESTIRGFMPTAGIOLISVLKYI 223
DB 151 -----YIQLVTFYTFPFGFIVQVAFGLERFGFVLEHLSQATIVGFLAAAVGIGIQOLKGL 205
QY 224 FGL-TIPSYGPGSIYVTFIDICKNLPHTNINSLIFALISGAEVLVKEELNARVHKIRF 282
DB 206 FGIDNFNNKTDLSFVSKSLTFSKNSAMHPYNIILIFSLCPIILFTFRGKNNKIKML 265
QY 283 PIPTEMIVVAVATAISGCC---MPKKYHMOIVGEIORGEPV-----SPVVSQW 330
DB 266 SHVAPLLSVSSAIKAVKINFELQVVDYKVAVIGPIKGSGLNPSLHQLTFDPSOVVGH- 324
QY 331 KDMIGTAFSLAIVSYINLAMGRILANKHGDVDSNOGMIALGCSNFGSFKIHVICCA 390
```







```

Db      365 YDPLDGNQELLALMGNTVCGVFRGAFAGSTALSRSAVSESTGKTQIAGLIGAIIVLIY 424
      419 MVLVGLVPLPKSVLGAALVAVLNKSLKQLTDPYILMRSKIDCCIMWVSFLSFLSL 478
      425 VLAIGFLVPLQSVLAALALGNLKMLOFAIGRLMRKDKYDCLIMTFTFTVLGL 484
      479 PYGAVGAVSVLVVVFQTOFRNGVALAQMMDIYVNPETKYRRAADIQIKITTCSP 538
      485 GGLASVAFQQLITVIFRQFPCSTLANIGRNIITKNKKDYDMDEPEBVKIFRCSP 544
      539 YFANSEIFPKQVY-----AKTVSLQELQO-----DEFNAP 568
      545 YFANIFEFRRKLIDAVGFSPRLIRKRNKLIRKIRLOKQGLQVTPKGEICTVDIKDS 604
      569 PTPNNQO-----TPANGTSVSY-ITFSPDSSSPAOSEPPASAEAPGEPMDLASVPPV 622
      605 DEELDNKQLEVLDPINTITDLPFHIMNDL-----PLNTEVRK-----I 644
      623 TPTHLLDMSGVSFVDLMGKALAKLSYTGKIGVAVFLVNIHAQVYNDISHGVEEDGS 682
      645 SLHSLILDFSAVSFLDVSSVRLGKSLTQEFIRIKVDYIYIGTDDFLERKNRYEFP-DGE 703
      683 LECKHYFPSTHDAVL 697
      704 VKSSIFFLTHDAVL 718

```

## RESULT 2

```

A54808
diastrophic dysplasia-associated sulfate transport protein - human
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 07-May-1999
C:Accession: A54808
R:Haesthacker, J.; de la Chapelle, A.; Mahrani, M.M.; Clines, G.; Reeve-Daly, M.P.; Daly,
der, E.S.
Cell 78, 1073-1087, 1994
A:Title: The diastrophic dysplasia gene encodes a novel sulfate transporter: positional
A:Reference number: A54808; MUID:95007757
A:Accession: A54808
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-739 <BA>
A:Cross-References: GB:U14528
C:Gene(s):
A:Gene: GDB:DTD; DTOST
A:Cross-References: GDB:125421; OMIM:222600
A:Map position: 5q32-q33.1
C:Superfamily: sulfate transport protein

```

```

Query Match      25.0%; Score 968; DB 2; Length 739;
Best Local Similarity 31.0%; Pred. No. 1,5e-66;
Matches 225; Conservative 139; Mismatches 285; Indels 76; Gaps 11;

```

```

      3 QRRPRYVDRRAVSLTLFDEFEKKDRTYP--VGKRLNNAFRCSSAKIKAVFGLLPVLS 60
      49 RYHRLILIRQ-----EKSDTNKEFVYIKLKQKCCSPAKAKMMILGFLVQ 97
      61 WLPKYIKYIIPDLGLSGSGSIQVPOGMAFALLANLAVNGLYSFFPLLTFFLGGV 120
      98 WLPRKIDKLNILGADVSLVGLVGLVLPQSIAYSLLAGQEPVYGLYSPFASIIYFFLIGTS 157
      121 HOMVPTGFVISILVGNIC---LQ-----LAPESKFOVNNMTNESYVDPAAM 165
      158 RHISVGFVGLCMIGETVYDRELQKAGYDNANSAAPSLGAVNSSTLLNHTSDICKCKSCY 217
      166 EBERLHVSAATLACLVAIQMGIGFMQGFVAIYLSSEFTRGFTWTAAGLIILSVLYKIFG 225
      218 ---AIWGSTVTFIAGVYVAMGCFQVSVYSLDALLSGFTVGSFTILTSOAKYLLG 274
      226 LTIIPSTGSGIVFTFIDICKNLPHNTNIALSLFALISGAFVLVYKLNARVMKIRFP 285

```

```

Db      275 LNLPRTNGVSLTIWTHIHFERNIKTNLCDLITSLCLLVLTPTKELNEHFKSKLAP 334
      286 TEMIVVAVVATISGGCKMPKXKMOIVGEIQRFPTPVSVVSQK---DMIGTAFSLAI 342
      335 TELVVVVAATLASHFCKLHENTNNSIAGHLPFGF---MPPKVEMULISVAVDAIATIS 391
      343 VSYVIMLAKGRTLANRGGVDVDSNQEMIALGCSNFGSPFKIHVICALSVTLAVDAGG 402
      392 IGFATIVLSJEMFAKKHGTAVKANQEVYALGFCNIILSPFHCTTSALAKTLVKESTGC 451
      403 KSOVASLGVSVYVITMLVGLITLPLPKSVLGAALVAVLNKSLKQLTDPYILMRSKID 462
      452 HTQLSGVAVTALLVLLVLAFLPFLYSLOKSVLGLVITIVMLRGLRFRDLPKMKMSISRD 511
      463 CCIMWVSFLSFPFLSPYGVAVGAVSVLVVVFQTOFRNGVALAQMMDIYVNPPTYNR 522
      512 TYIMPTMILSALLSTEIGLVGCSICVILIRTKPKSSLLGLVEESVFESVATYKN 571
      523 AADIQIKITTCSPLYFANSEIFRQKVIKATVYSLQELQODEFENAPPTDPNNQTPANGT 582
      572 LQTKPGIKIFRFVAPLYIINKECFSALYKQTVNPILIKVAMKAKRKIKERVTVLGI 631
      583 SVSYITFSPDSSSPAOSEPPASAEAPGEPMDLASVPPVTFHTLLDMSGVSFVDLMGI 642
      632 Q-----DEMVSQLSHP-----LELHTVIDCSALQFEDPTAGI 664
      643 KALAKLSYTGKIGVAVFLVNIHAQVYNDISHGVEEDGSLECKH-----VFPSTHDAVL 697
      665 HTLKEVRRDYEALGIVGLAQCNPYRDSLTITNEY-----CKKEENLTFISVEAMA 717
      698 FAQAN 702
      718 FAEVS 722

```

## RESULT 3

```

A49994
sulfate transporter, canalicular - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Jan-1995 #sequence_revision 20-Jan-1995 #text_change 24-Sep-1999
C:Accession: A49994
R:Bissig, M.; Hagenbuch, B.; Stieger, B.; Koller, T.; Meier, P.J.
J. Biol. Chem. 269, 3017-3021, 1994
A:Title: Functional expression cloning of the canalicular sulfate transport system of
A:Reference number: A49994; MUID:94132077
A:Accession: A49994
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-703 <BIS>
A:Cross-References: GB:L23413; NID:9431452; PIDN:AAAI7545.1; PID:9431453
C:Superfamily: sulfate transport protein
C:Keywords: glycoprotein; transmembrane protein

```

```

Query Match      22.0%; Score 853; DB 2; Length 703;
Best Local Similarity 29.3%; Pred. No. 1e-57;
Matches 201; Conservative 143; Mismatches 304; Indels 38; Gaps 8;

```

```

      36 KLNNAFRCSSAKIKAVFGLLPVLSWLPYKIKDYIIPDLGLSGSGSIQVPOGMAFALL 95
      34 RLKSCCTGSPCAQALVQGLFPIRMLPQRLKLEYLAGDMSGVLGIILVPOALISLL 93
      96 ANLPVANGLYSFFPLTFFFLGVAQWVPGFAVISILVGNIC---LQLA----- 143
      94 AGIQPIYSLTSPFANLIFLMGTSNHWVGFSLCLWVGQVYDRELQAGDPSQDSL 153
      144 -PESKFOVENN-ATNESYVDPAAMEARLHVSAATLACLTATIIQMGIGFMQGFVAIYLS 201
      154 GGNNDNSTNNATLVLVGLQDGRDCHAIRIATALTALMAGLVQVLMGIRLGFVSTYLSQ 213
      202 SFIRGFTMAAGIILSVLKYIFGLTIPSYTSGSVIVFPIIDICKNLPHNTNIALSLFALI 261
      214 PLDDGFAMGASVITLTSQAKHLGLVRIPIRHQGVNIHTWLSLQWVGANLQDVVTSAV 273

```



Qy	262	SGALVLVYKELMARMYMKHRIEPIPTMTIMVYVATASGCGCKPKKYIHHQIVGELORCEPT	321
Db	274	CLAVLLLAKKELSDRYRHYRLKVPVPELTLLIVATIASHFQGLHTRFGSSVAGNIPTGEVA	333
Qy	322	PVSPVVSQMDMIGTAFSLAIVSYVNLAMGRTLANKHGVDYDSNOEMIALGCSNFCFSF	381
Db	334	POIDPRKIMSVALLDAMSLAVGSASFISISLAEMFARSHQSYVSANOELLAVGCCVLPAF	393
Qy	382	FKIHVICCALSVTLAVDAGAGKSOVASLCLSVYVMITMLVIGIYLYPLKRSYLGALLAVN	441
Db	394	FHCFTSAALSKTLVYKATGCGCOTOLSSVYSAVWLVLLVTLVLAPLRPHDQRCVLACTIIVS	453
Qy	442	LKNSLQLOTPBYLYMRKSTKIDCCIMVVSFLSEPSFLPGVAGVAFSVLYVPEOTOPRN	501
Db	454	LRGALRKVKKDLPOLMKRSLSPDALWYATATCYLVSIENGLAGVFFSLISLAGTQOPR	513
Qy	502	GVALAQVMDTIDVYNPKTYNRADIODIGIKIITYCSPLYFANSEIFROKVIKTYVSLQELQ	561
Db	514	AALLARIGDSTVEEDAAEFEGGLPPEVRYFRFPTGLYANKDF----LRSTYSLGLD	569
Qy	562	QDEFNAPPTDPNNQPTANQTSVSYITTESPDSSPRAQSEPPASALAPGEPSMLASVPPF	621
Db	570	AGY--SARRKORGTVEGVSNRSVLDRKDLQSVS-----GGGLV-VPLA	610
Qy	622	VTFHTLLIDSGVSFYVDLMKIKALAKLSHYGKIGVLYIHAQVYNDISHGCVF--ED	680
Db	611	FGFHTVYIDCAPLFLFDVAGMATLKLDRKNRYALDITLLLACCSPSVRDLTKKGGLGED	670
Qy	661	GSLECKHYVPSIHDAVLYFQAQANRDV	706
Db	671	QGTAEELLFVPSVSHAVETACARREEL	696

```

RESULT      4
T27820
hypothetical protein ZK287.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T27820
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27820
A:Status: preliminary; translated from CB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-611 <FWL>
A:Cross-references: EMBL:Z70757; PIDN:CA94798.1; GSPDB:GN00023; CESP:ZK287.2
A:Experimental source: clone ZK287
C:Genetics:
A:Gene: CESP:ZK287.2
A:Map position: 5
A:introns: 3/3, 102/1, 438/3
C:Superfamily: sulfate transport protein

```

Query Match	20.28	Score 781	DB 2	Length 611
Best Local Similarity	28.38	Pred. No. 3e-52		
Matches 197; Conservative 138; Mismatches 262;			Indels 98;	Gaps 12

```

0Y      18 TLIDDEFKKRDRTPVGEKLLBNARCSSAKITANVEGLLRLVPLSLPKRYTKIDVTIIPDLG  77
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      9 TTTDDDIYEKREKCSFRKKLCK-----YILELMLPVQMKMDHFGDVA  53

0Y      78 GLSGSIOVPGGMAPALLANPAVNGLYSSFFLLTFPLGVMQMPGFMAYTSLVGN  137
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      54 GLTGTIMVPGGMAYASLAGVPVYAGMSSEFASTTMEFGTARHHSIGFVAASMMVGA  113

0Y      138 ICLGLAPEKSPROVENNATNESVDTAAEAEARLIVASTACTGTAITOMGSGFMQFQVAI  197
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      114 ARRLRAPD--IPISNSSIINSVPLGGEYDPLVFTSLALLVLGVSVDITMGILRGELTT  171

0Y      158 YLSESFIRGFMTAAGLOLLISVLKYIFGLTIPSTYSGEISVFFPIDCKNLPHTINASLI  257

```

```

Db      172 YLSSOLVSGFTTGAANVHFTSOLNKVPGILTRPHEIGIAVMKRYKDMIMSLGVSFVALG 231
Oy      258 FALISGAEFLVLYKELNARYHKTI----RPRIPTEMIVVVVAITALSOGCKMPKRYHMOIV 312
Db      232 ISI-----FGLIFLIDLGRITVINPIYKFKSPPIRPLELILVIFGLIVISMIPNLDAEYHNVTV 287
Oy      313 GEIQGRPPTPVSPVYSOMKDMIGTAFSLAIVSVINLAAGRILANKHGDVDSNOEMAL 372
Db      288 YEIPRGPPLSPRIPLNLFPLALLSDAIPRIAVAVCMFVMSMGKLFPAKKHKRYKTATGTELAI 347
Oy      373 GCSNPFSGFEKTHVYICCALSTVLAVDAGAGSOVASLCSLVMMITMLVLGIYLPRLKS 432
Db      348 GIASHLSSEFPYVYVVGASLSRSSVCESGANTDYLITPSEFLITVILILGFLERPLMC 407
Oy      433 VLGALIAVNLKNSLKQLTDPYLYLMRKSJLDCCIWVVSFLSFFLSLPRYGAVAVGAFSVLV 492
Db      408 ILACIVAIYSLKSLPMQYKELPRILYIRISKYDFALIMLVACSLTFTPDVTTGLVISTLSFSLYT 467
Oy      493 VYFQIQFNGNGLAQNWDITIVNPKRYKNAQOI--QGIKITIYCSPLYRANSEIRAKOVI 551
Db      468 LVLRQWQ---SEFTYLAHDE---TPR-----QWVPNVAIKYFAGSLIHANVTAR---- 511
Oy      552 AKTVSLQELQDFENAPRPTDPNNQRPANGTSVYITFSPDSSPAQSEBPASAEAPGEP 611
Db      512 -----QDDM-----GEA 518
Oy      612 SDMLAVPPVFTFHTLLIDMSGVSEFYVDLMGIKALAKLSTYGGIGVKEFLVNIHQAQVND 671
Db      519 IGLRPEEDPLIDERITIIILDASSVAFIDIMGVADLRQVFKDALKTGIVHVEYCGLPREDVLSV 578
Oy      672 ISHGVEFEDGSLCKHVPSPHDAVLEFAQANADV 706
Db      579 LSNDENFL-SVAPPESTPEPST--DSCILSEVHQOVS 611

```

RESULT 5  
T23628  
hypothetical protein K12G11.1 - *Caenorhabditis elegans*  
C.Species: *Caenorhabditis elegans*  
C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C.Accession: T23628  
R.Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A.Reference number: Z19774  
A.Accession: T23628  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-749 <WtL>  
A.Cross-References: EMBL:Z81570; PIDN:CAB04606.1; GSPDB:GNO0023; CESP:K12G11.1  
A.Experimental source: clone K12G11  
C.Genetics:  
A.Gene: CESP:K12G11.1  
A.Map position: 5  
A.Introns: 48/3; 128/1; 165/1; 220/3; 407/3; 471/3; 503/3; 539/2; 585/2; 629/3; 669/3  
C.Superfamily: sulfate transport protein

Query Match	20.18;	Score 776.5;	DB 2;	Length 749;
-------------	--------	--------------	-------	-------------

Matches 206; Conservative 142; Mismatches 311; Indels 71; Gaps 13;

```

0Y 2 SOPRRRYV--VDRAAYSLTLDDEEKKEDRTYVPYGEKIRRNAPRCS$AK-----47
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 31 SEPNNRRSVFVGRGAMNOAQDEK---DYNNP---HLENLTKQAKQFVRRFEPTSF 84

0Y 48 --IKAVNGELPVLSMLPKYTKIDYITPDLLGSGSIOYPOGMATALLANPAVNGLY 105
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 85 FALKLEIFDLDLPILKMPFEPYKMTDLSIDITIGITVGMOPVQISIAVALLAKOAINGLY 144

0Y 106 S$EPFLPLYFELG$YH$V$M$PCTFAVISTIVNIGIC$Q$LA$B$SK$FQ$VFN$NAT$N$SY$D$T$A$M 165
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 145 $L$P$P$P$L$Y$M$L$G$S$R$H$A$S$L$T$F$A$V$S$T$M$L$O$V$E$K$L$A$A$P$D$Y$-P$S$F$E$D$D$D$V$L 2020

```

Oy 166 EAEERHVSNTAACLTAIITQMGLGFMQGFVAIYIYSESTIRGEMTAAGIQLISVUKITFG 225  
 Db 203 PSP-TEVSCALITTMGLLFLVWGVLRLQFLTTYLSDVYIAGFVGSVHVLVSQKLTLLG 261  
 Oy 226 LT-IPSTYGPGSIVFTTFIDICKNLPHTNIASLIFALISGAFVLUKELNARYMH---KIR 281  
 Db 252 IRLGLRHSRGPFLFPHNLVDLWSTIRANPVSGISIVSIIVLHCKEETINPFIKKTKSN 321  
 Oy 282 PPIPTETIVVVVATPISGCKMKPKYHMQIVGELTGRCPTPVPSPVSOQMDIGTAFLA 341  
 Db 322 IPIPELVIYILSTVFVAVTGVDTEAKVQVYNNKIPVGNPSLSPLYILPQVLPAISIT 381  
 Oy 342 IYSYIYNIAMGRTLANKGIVDYSNQEMIALGCSNFSGSFKIHVICCALSVTLAVDAG 401  
 Db 382 VYSISVWLSISMLAKRYNELDSGDELFAISISSFTPIPSNCSLRTLEAVAG 441  
 Oy 402 GKSQVASCVSIVNMTMLVIGIYDPLPKSVLGLIVNKNLSKJOLDTPYYILMRSKL 461  
 Db 442 CTTOIISTFSSILIVSVYIFGLTLETLPMALSAITICVALQGMFRKPADLIDLKASKI 501  
 Oy 462 DCCIMVWSFLSFLSLPYGAVGAVAFSVLVVFEQTGRNGYALAOVMDTDIYVNPKN 521  
 Db 502 DFTIMVWCVSTIILDVSTGLISVCFALFTTILREQYPKMHLLASVGTODFRAEKYG 561  
 Oy 522 RAQDIQGIKITTCGPLYFANSEIFRQVIKYVSLQE-----LQDDFENAPPTDPNN 575  
 Db 562 ETVYVFKGICIFEFDPALPLFHNVCEFKSIEKAYTEMOKSHEFYVLRERETILNKLDS 621  
 Oy 576 QTPANG----TSVSI-I-TPSPSSPAQSEPPASAEAGPESDMLASPPVYTFHTLLD 630  
 Db 622 DESIDGKMFQTAOSTLNTNSPD-----ILSRHEVID 652  
 Oy 631 MSGVSFVDMGKALAKLSTYGIKGVKFLVNIHAQVYNDISHGVFEDSLECKHYFP 690  
 Db 653 CSGFIFIDLMGSAKELFSDMRKGLIYFANNAKAPVREMPKCHFFNPFYSKE--NRY 710  
 Oy 691 SIHDAVLEPAQ 700  
 Db 711 TMRDATSIAR 720  
 RESULT 6  
 T23629  
 hypothetical protein K12G11.2 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T23629  
 R:Kershaw, J.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19774  
 A:Accession: T23629  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1700 <WIL>  
 A:Cross-references: EMBL:Z81570, PIDN:CAB04607.1; GSPDB:GN00023; CESP:K12G11.2  
 A:Experimental source: clone K12G11  
 C:Genetics:  
 A:Gene: CESP:K12G11.2  
 A:Map position: 5  
 A:Initons: 26/3; 52/; 107/1; 197/3; 253/3; 307/3; 335/2; 480/3; 516/2; 562/2; 608/3; 66/3  
 A:Superfamily: sulfate transport protein

	Query Match	19.9%	Score 768	DB 2:	Length 700;
	Best Local Similarity	28.3%	Pred. NO. 3.7e-51;		
	Matches 199;	Conservative 141;	Mismatches 322;	Indels 40;	Gaps 11.
Oy	12 RAASLTLEFDDFEKKDRPYVGEKLRNAPRCS-----AKIKAVFCGLPVLNL 62				
	: : : : :				
	: : : : :				
Db	21 RAPMQLVEDEYGGOKRKDGGRKKKSTKASHRYVPETSVTFKFTFLMLPFIEGWL 80				
yy	63 PKYIKDIYIPDLGLGSGSIQVPOGAFAALLANLPVNGLSYSSFPPLTLTFFLGVAHQ 122				

```

Db      81 PXYDMKNLSLTVGVGTTGVQIOPGIATVALLSNODPIVGLXTSYTIPFLLTFFGTSKH 140
QY      123 MWGTFPAVISILVGNICQLAPESKFQVFNFNATNESYVDTAAMEARLHVSATLACTAI 182
Db      141 ASLGTFAVVALMTG-----LAVEREAFIPSDMLNSTLLPGDEALSPLEVSACALVGL 195
QY      183 IOMGFGMOFGVAVIYLSSEFIRGFMPTAAGLOILSLVLYKINGLP-IRSYTGPSSIVTFE 241
Db      196 IOPLMGVRRLOPITLYLSQOLIAGFTTGSAAVVAISQREKELPGLGLVAKHSPGLIRNV 255
QY      242 IDICKNLPHPTNIASLIFALISGAFILVKE-LNABYMHKIR--PPIPTEMIVVAVATAIS 298
Db      256 YDFVNLPRKAMCCASISATIMILLHCKEYINPIMKRKMKSNIPILMELVAIISTIFV 315
QY      299 GGCKMPKRYKHYOIVEIORGFPTPVSPVSOXKMDIGTAFSAIYSYVNIAMGTILANK 358
Db      316 ALIDANELNVAIVYKIPGLPELSLPNPENLIPRVLDPDISIAVVAHVHLSLSMLKK 375
QY      359 HGVDYDSNOEMTALCSCNFFGSFRIHYICALSVTLAVDAGSKSOVASLCSLVMT 418
Db      376 YEYEIDAGEIYALSFETALIGSGFFPTPTISIGRTWGVESGVKTYATFSCLELVSY 435
QY      419 MVLGLIYLPRLKSVUGALIANVNLKNSLKOLDTPYRLMRKSLDDCINWVSLSFPLSL 478
Db      436 SLYERFRETTPCVALSALITVATLAKSMLKRLDLKGIMKLSIDCCIMWVAFAVAVLDV 495
QY      479 PLYGVAVGAVSFVLVVFQTOFRNGYALQAVMDTJIVYVPKTYNRADODJIGIKIITYCSPL 538
Db      496 SEGILLIAFFALFTTLRQYPKMHLANVKRDTDFSDIOQOETIFPKGICIFEDAPL 555
QY      539 YFANSEIFROKIATVLSIQELODFENAPPTDPNNNOTPANGTSVSVITSPDSSPAQ 598
Db      556 LEHANECKR-KCIEKVYDEMKKSEF-----NFKEPNAGKGSFTEGMRHIAPI 605
QY      599 SEPPASAEAPGEPSSMLASVPPEVTFHTLIDMGVSEFUDLMGIKALKLSSTYKIGV 658
Db      606 TEIPLH---PGINRD-----PILPRH-FVIDCSGFTTIDLMGVSALKEVPSDLKRRVQ 655
QY      659 VELVNIHMOYVNDISHGVFEDGSLCKHVPSPHDAVLAQ 700
Db      656 YFASSTKVAVREMPKCSFFDFVSK--NFYPTLADAGIAR 695

```

RESULT 7  
T16077  
hypothetical protein F14D12.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C:Accession: T16077  
R:Minx, P.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid F14D12.  
A:Reference number: Z18457  
A:Accession: T16077  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-650 <MIN>  
A:Cross-references: EMBL:U01021, NID:g1006679; PID:g1006682; PIDN:AAA8235.1; CESP:F14D12.5  
C:Genetics:  
A:Gene: CESP:F14D12.5  
A:Introns: 5/3; 53/2; 77/1; 159/3; 181/2; 215/3; 270/3; 321/1; 443/3; 479/2; 503/3; 507/2  
C:Superfamily: sulfate transport protein

```
Query Match Similarity    24.1%; Score 631; DB 2; Length 650;
Best Local Similarity     24.1%; Pred No 1.2e-40;
Matches 179; Conservative 145; Mismatches 290; Indels 128; Gaps 17.
```

[illegible]

RESULT 8  
T25751  
hypothetical protein F41D9.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T25751  
R:Fulton, B.  
submitted to the EMBL Data Library, August 1996  
A:Description: The sequence of *C. elegans* cosmid F41D9.  
A:Reference number: Z20081  
A:Accession: T25751  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Residue type: DNA  
A:Molecule type: DNA  
A:Residues: 1-758 <F41D>  
A:Cross-references: EMBL:U67954; PIDN:AB852608.1; GSPDB:GN00028; CESP:F41D9.5  
A:Experimental source: strain Bristol N2; clone F41D9  
C:Genetics:  
A:Gene: CESP:F41D9.5  
A:Map position: X  
A:Introns: 5/3, 74/1, 111/1, 165/2, 206/3, 261/3, 313/3, 370/3, 466/3, 502/2, 551/3, 585/3  
C:Superfamily: sulfate transport protein

Query Match	15.28; Score 588; DB 2; Length 758;
-------------	-------------------------------------

Best Local Similarity 24.68; Pred. No. 3.1e-37;  
Matches 175; Conservative 141; Mismatches 295;  
Indels 100; Gaps 18;

```

0Y 35 EKLNAFRCSSAKIKAVFGLPVLMLPYUKYIKDYIIPLLGLSGSGSVQOGAFL 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 22 OKLR--YACSRSKCIHSLSLFLPIITMLPKYDKMSSHFFGLSGCLTMAYSVQGIALAS 79
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 95 LANLPAVNGLYSSFFPLLTFFELGCVHOMVGTFAVISIL----VGNICLOLA----- 143
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 ITGVPPVGLTYATJAFPSFLYIFFGTSKHNALGCAVLSLMTHALIEKVMIRATSYNATA 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 144 -----PEKFOYFNMT-----NE-SYVDRAAME-----AERLHVSEATLA 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 YVNHTELDLDEKNEFALISMTLMQILNETSVBEVTEMMTEGVTYKQIHNVAATTII 199
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 178 CLTAIIOMGLCFOMQFGVAIYLSSEFIRGEMTAAGLOILISVLKYIFGLIPSYTPGSI 237
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 FLAVGIQVFMGVFLQYLTLSLFSQVMSGFVGGGIHVFAQIGNMLGIELPRSGGYL 259
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 238 VFTFDICKNLPHNNIASLIFALISGAFVLVKLARKMYHK-IRPFIPEMLVYVYATA 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 YRIMDLEVNEDNHVIFVYCISLSEFLFVGEKRYLAPMLNSAFNVPPELVLTT--- 315
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 297 ISGCKMKPKKTHMOIVGEIQRGFTPPSPVVSQMKDMIGTAFSLATSVYINILAMGRTLA 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 -----HYL-----NRPPSLPRRDLRIHGLMAALATAIYAHITVAKVE 359
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 357 NKHGIVDSNOEMTALGCSNFGSFEFKIHIYICALSVTLAVDAGKSOVASICVSLVYM 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 KRYKXKINHGOELALGFVGLSEFFPVFPVPTSGFASVVGAAVGGSTOLTCLCFSSLATL 419
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 417 ITMLVLGIYILPKRSVYLGALIVANLKNLSKOLDPPYLRKSKSLDCCIVWFSFLSFL 476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 SVIILCIGALEYLTQCLISAMITIFAQGMLEKBEKLSIMPVKRIIDPTIYMSFELTYC 479
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 477 SLPGVAVGAFSVLVVPOTQFNGYALAAQVMDTIVNPKTYNNRAODIOG-IKIITYC 535
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 DMGEGLLMAIOGFAVLTIIIFQREKMHFLSDDTENTYKETKRKD-LEKIOGVNCJFRMD 538
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 536 SPLFANSEIYRQVIAKTVSLOLOODEFNAAPPTDNNNQTANGTSVAYIT---FSPD 592
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 APLIFTSSDR-----THSVOCYKWEK-----KSEFVITIEQMSND 577
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 593 SSSPAOSEPPASAE---APGEPSDMLASVPEYTFHTLLIDMSGVSFVDMGIKALAKLS 649
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 RSADIFDSKLSARRMRKRDQKSENRC-----LVIDCGEFPYVDYLGSLTKSVY 628
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 650 STYKGIQVFLVNIHMOYNDIHGCVFEDGSLJECKHVPSTHDAFLFMO 700
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 629 VDLQAAIGQCFV--VQKSDLKKLRATDFEYVDESK-VENKYGDAVKAKE 677

```

RESULT 9  
T26165  
hypothetical protein W04G3.6 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T26165  
R:Renard, N.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z20163  
A:Accession: T26165  
A:Status: preliminary; translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-380 <WLL>  
A:Cross-references: EMBL:Z68014; PTDN:CAA92028.1; GSPBP:GN00028; CESP:W04G3.6  
C:Genetics:  
A:Experimental source: clone W04G3  
A:Gene: CESP:W04G3.6  
A:Map position: X  
A:Introns: 46/1; 83/1; 135/3; 153/3; 190/3; 244/3; 272/2

A; Introns: 46/1; 83/1; 135/3; 153/3; 190/3; 244/3; 272/2



```

Oy 173 SATLACIAIIOMGFMQGFVAIYLSSEFIRGFMTAAGIOIISLVTUKYFGLIPBXT 232
Db 167 AFTATFEGVLEASLGIRLGFIDFELSHATIVGMSGAATVVSIOQLKGIFG--KHFT 224
Oy 233 GPGSIVTFEFDICKNLPHNTIASLIFALISGAFVLVAKELNARVNHKTRFPIPTMIYVV 292
Db 225 DSTDIVISVMRSVFQOTHMRMS---GVLCGCFLEFL--LSIRY-----FSTKKRKFVW 274
Oy 293 VATA-----ISGCKM---PKYHMOIVGEIQRGEPFV-----SPVSOXKMDMIG 335
Db 275 AAMPRTSVIIIGSLIVYFTHAERHGVQVIGDLKGL-NPLSGDILFTSPYMS--TAVK 330
Oy 336 TAFSLAIVSYIYNIAMGRTLANKHGVDVDSNOEMALDCSNFSGSEFKIHYICCALSTL 395
Db 331 TGLITGIIALBGLIAGVSGFAFKNYNTDGNEKEMIAFPMNNTVGSFTCYLTGTGFSRSA 390
Oy 396 AVDGAGCKSOVASLDCVSLVMITMLVGLIYLPPLPSVIGALIANLKNLSKOLDPYYL 455
Db 391 VYNMNGCKTAMSNIMYALVAMFTLLFLPLRFHYPLVYLSAI--ISAMGLIDYQAIHL 449
Oy 456 WRKSKLDCCIMVVSFLSSEFLSLPYGVAVAFSVLVVYFQTFRNGYALAOVMDTIV 515
Db 450 WKVDFEFLVCMASAVGVFGSVEIGLVAVVAISARLLFEVSRPKTVKGNIPMSIYR 509
Oy 516 NPKTYNRADLOGIKIITIIYCSPLVFANSEIFROKIYATVSIQELQOOFENAPPTDPNN 575
Db 510 NTEOIPSSRTVPGLIILEIDAIYFANASYLREIRIRW-----IDEEEE 553
Oy 576 QTPANG--TSVSITTFSPSSSPAQSEPPASAEAPGEPESDMLASVPFVTFHTLLIDMSGV 634
Db 554 RVKSGGESLQYI-----IIDMSAV 573
Oy 635 SFVDLMGIKALAKLSSTYTGKIGVKEFLVNIHAQVYNDISHGVFEDGSLCKHVPSPSHD 694
Db 574 GNIDPSGISMMVEIKKVIDRALKLVLSNPKGVEVKKLTRSKFIOD-HLKGEMFLTVGE 632
Oy 695 AV 696
Db 633 AV 634

```

RESULT 12  
T49069  
sulfate transporter (ATST1) - Arabidopsis thaliana  
N:Alternate names: protein F4F15.10  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T49069  
R:Alvarez, J.P.; Claphault, G.; Cottet, A.; Maché, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25015  
A:Accession: T49069  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-658 <ALC>  
A:Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.10  
A:Experimental source: cultivar Columbia; BAC clone F4F15  
C:Genetics:  
A:Gene: ATSP:F4F15.10  
A:Map position: 3  
A:Introns: 116/1; 185/2; 242/3; 264/3; 302/3; 341/3; 384/1; 479/3; 543/2; 572/1; 597/3  
C:Superfamily: sulfate transport protein

	Query Match	14.0%;	Score	543.5;	DB	2;	Length	658;
	Best Local Similarity	24.9%;	Pred.	No.	7e-34;			
	Matches	180;	Conservative	127;	Mismatches	274;	Indels	141;
					Gaps	23;		
Oy	4	PRPRTVDRAAVSL--TLF-DDEFKSDRTIVGKKLNATFCSSAKIKAVFGL---	Lp	57				
		:   :	:	:	:	:	:	:
Dd	25	FQPPFLSISLQSVAKETIPDPDFRQ-----FKNNASR-----KFPLSGKYFLP		69				
Oy	58	VLSWLPKKYKIDYIIPDLIGLGSGSIQVPQGMAFALLANLPNAVNGLYSEFPPLLTYFEYL		117				

```

Db 70 IFENAPRNLK-FEKSDLIAGTTIASLAIPOGISIAKLANLPRILIGLSSFFPLVAVL 128
QY 118 GGVHOMVPGTFAVISILVGNICLOLAPESKROVENNATNESYVDTAAMEAER----LHV 172
Db 129 GSSRDLAVGTVAVASLITLG--AMLSTKE-----VDAEKDPKLYLHL 166
QY 173 SATLACLATLAIOMLGEPNOGFVALIYLSSESPTRGPMTAAGLOILLSVLKYIFGLTIPSYT 232
Db 167 AFTATFAGVLEASIGIFRLGIFDELISHATVGMGAATVVSLOOLKJLGF--KHFT 224
QY 233 GPGSIVFFIFIDCKMLPHTNIALIFALISGAFVLVYKELNARYMHKIRPIPTEMIYV 292
Db 225 DSTDVISVMSRVSFQSTHEMRMS---GYLGGCFLLFL--LSRY-----FSLKRRKFFW 274
QY 293 VATA-----ISGGKM---PKYHMQIVGETIQGEPPIV-----SPVSOQKMDIG 335
Db 275 AAMAPLTVIILGSLLYVFTHAERHGVQVIGDLKGL-NPLSGSDLIFTSPYMS--TAVK 330
QY 336 TAFSLAIYSVINIAMGRTLANKHGVDVDSNOEMIALCGSNFFGSEFFIHYICALSLTL 395
Db 331 TGLITGIILALBEGVAVGHSFAFENKYNIDGNKEMTAFGMNIVSGSFCTYLLTGTFPSRSA 390
QY 396 AVDGAGKSOVASLDCVSLVVMITMLVLGILYLPULRSVAGALIANLKNLSKOLDPPYL 455
Db 391 VYNNNGCMTAFMSNIYMAIAVMFTLLFLPLRPYPLVYLSAI--ISAMIGLDYQAAIHL 449
QY 456 WRKSKLDDCCIMVVSFLSSEFLSLPYGVAVGAFSVLVYVFOTOPNGSALAAVMDTIYV 515
Db 450 WKVDFEFLCYMSAAYGVGVFSGVEIGLVAAVAISIAIRLLFEYSRPKTAVKGNIPMSIYR 509
QY 516 NPKTYNRARDIOGIIITYYCSPLYFANSEIPROKIAKTAVSLQELQOPEMNAPRDPNNN 575
Db 510 NTEOYPPSSRTYFGLILLEIDAPLYFANASYLERIIR-----IDEBEE 553
QY 576 QTPANG-TSVSYITFSPDSSSPAQSEPPASAEAPPEPSPDMLASVPFVYFHTLLIDMSGV 634
Db 554 RYKQSGESSLQI-----LIDMSAV 573
QY 635 SFVDIMGIKALAKLSSTYGKIGVKVFLVNIHAQVYNDISHGCVFEDGSLCKCHVPPSIHD 694
Db 574 GNIDISGISMVWEIKVIDRRLAKLVLSNPKEGVYKKLTRSKFID-HLGKEMPLFVGE 632
QY 695 AV 696
Db 633 AV 634

```

```

RESULT 13
T51161
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 02-Sep-2000
C:Accession: T51161
R:Comella, P.; Wu, H.-J.; Landie, M.; Berger, C.; Cooke, R.; Delseny, M.; Girellot, F.
Plant Mol. Biol. 41, 687-700, 1999
A:Title: Fine sequence analysis of 60 kb around the Arabidopsis thaliana AtEm1 locus
A:Reference number: 224835; MUID:20108326
A:Accession: T51161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-703 <COM>
A:Cross-references: EMBL:AF049236; PIDN:AMC1417.1
C:Genetics:
A:Map position: 3
A:Introns: 161/1; 230/2; 287/3; 309/3; 347/3; 386/3; 429/1; 524/3; 588/2; 617/1; 642/2
C:Superfamily: sulfate transport protein

```

Query Match	14.0%	Score	543.5	DB 2	Length	703
Best Local Similarity	24.9%	Pred. No.	7.7e-34			
Matches 180	Conservative	127	Mismatches	274	Indels	141
					Gaps	23

```

0Y 4 PRPRVVDRAVYSL-TLF--DDEEKKDRTRYPGEXETLRNMFROCSAKIKAVNGL---LP 57
Db 70 POPOFFLSLOYKVEETLEFPDDRÖ-----FKÖNÖSR-----KEVLGLXYFLP 114
0Y 58 VLSWLPKKYKINDYIIPDLGLSGSGSIQVDPGMAFALLANLPAVNGLYSSFFPLTYFEL 117
Db 115 IFEMAPRYNLK-FPKSDLIAGITIASLAIPOGISYAKALANLPLILGYSSEFPPLVAVYL 173
0Y 118 GGVHOMVPGTAVVISILVGNICLLOLABESKRYOVFNATNESYVDTAEMEER----LHV 172
Db 174 GSSRDLAVGTAVVASLTLG--AMLSKE-----VDAEKDPKLYLHL 211
0Y 173 SATLACLAATIIOMGEGMÖGEFNAVITYLESSEPIGEFMTAAGLÖLISVXKIFPDLTPST 232
Db 212 AFTATFPGVLEASLGFIRLGFIDFISHATYVGEFGAGATVLSLOÖKIFEL--KHFT 269
0Y 233 GPGSIVFPIFDICNKLPHNTIASLIFALISGAFVLVKEANARMHKIRPIETEMIVVY 292
Db 270 DSTDVISWRSVFSGOTHEWRSES---GLÖCGFLFEL--LSTRY-----FSIKKPFVWY 319
0Y 293 VATA-----ISGCKM---PKKYHMOIYGEIÖGFPTPV-----SPVVSÖMKDMIG 335
Db 320 AAMPRTSVILGSLLYVFTHAERHGVOVIGDLKKGL-NPLSGSDLIFTSPYMS---TAVK 375
0Y 336 TAFSLATSYVINLAMERTLANKIGYVDVDSÖEKMALGCSNFGSGFPFKIIVICALSIVL 395
Db 376 TGLITGIILAEVAGVAGSFAMFKYNNIDGKEKELIAGEMNIVGFSFLCTLTTPESRSA 435
0Y 396 AVDGAGCKSOVASLCSVLVWITMLVIGIYLPRLPKVLCALLAVNLKSLKÖLQUTDPYVL 455
Db 436 VYNAGCGCTAMSNIVMAAVWETLLEFLPLFHTPLVYLSAII-ISMGLIDYÖAATIH 494
0Y 456 WRKSKLDCCIIWVSFLLSFFELSLPRLGYAVAGVAVSVLVVVEFOTGRNGYALAOYMDTDIYV 515
Db 495 WKVVKFDELVCNMSAYGVGVFSGVIEGLVAVVAISIRALLFLFVRPKTAVYKGINPSMIR 554
0Y 516 NPKYNNRÄÖDÖGKIITTYGCPLEFANSELPÖKVYIAKTYSDLOELODEPNAPPTPPNNN 575
Db 555 NTEÖPPSSRYPLLEIDAPILFAYNASYLREIIRW-----IDEEB 598
0Y 576 ÖTPANG-TSVSYITFSPDSSSPAÖSEPPASAEAGPEPSDMLASVPPTVPTHTLILMSGY 634
Db 599 RVKÖSGESLOY-----ILMSAV 618
0Y 635 SFVDLMGICALAKLSSTYGIKGVKVLVNIHAÖYNDISHGVFEEDGSELECKHVPFSIH 694
Db 619 GNIDTSGISMMVEIKKVIDRALKTLVSNPKGEVYVKKLTRSKFIDG-HLCKEMWFLTVE 677
0Y 695 AV 696
Db 678 AV 679

```

RESULT 14  
T32945  
hypothetical protein W01B11.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #ext\_change 17-Mar-2000  
C:Accession: T32945  
R:Bradshaw, H.; Graves, T.; Blalr, T.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid W01B11.  
A:Reference number: Z21250  
A:Accession: T32945  
A:Status: preliminary: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-809 <BRA>  
A:Cross-references: EMBL:AF043704; PIDN:AAB97595.1; GSPDB:GN00019; CESP:W01B11.2  
A:Experimental source: strain Bristol N2; clone W01B11  
C:Genetics:  
A:Gene: CESP:W01B11.2  
A:Map position: 1  
A:Introns: 15/2; 38/3; 114/1; 269/3; 420/1; 497/1; 518/2; 580/1; 613/1; 648/1; 684/1; 72

C; Superfamily: sulfate transport protein

Query Match	13.8%;	Score 532.5;	DB 2;	Length 809;
Best Local Similarity	23.8%;	Pred. No. 6.6e-33;		
Matches 178;	Conservative 118;	Mismatches 274;	Indels 177;	Gaps 19;

[illegible]

RESULT 15  
T48902  
sulfate transporter AST12 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T48902  
R:Takahashi, H.; Sasakura, N.; Kimura, A.; Watanabe, A.; Saito, K.  
Plant Physiol. 121, 686, 1999  
A:Title: Identification of two leaf-specific sulfate transporters in Arabidopsis thaliana  
A:Reference number: 225001  
A:Accession: T48902  
A:Status: preliminary; translated from GR/EMBL/DBJ  
A:Molecule type: DNA







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:02:14 ; Search time 14.18 Seconds  
(without alignments)  
1194.991 Million cell updates/sec

Title: US-09-749-589-2  
Perfect score: 3869  
Sequence: 1 MSQPRPRRYVDRRAVSLTLF.....WDLQDFGSMFRAETLAL 753

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1120	28.9	764	1 US-08-424-567-2	Sequence 2, Appli
2	1120	28.9	764	2 US-08-711-928-2	Sequence 2, Appli
3	1120	28.9	764	4 US-09-184-937-2	Sequence 2, Appli
4	123.5	3.2	470	1 US-08-471-436-2	Sequence 2, Appli
5	123.5	3.2	470	2 US-08-894-840-2	Sequence 2, Appli
6	123.5	3.2	470	3 US-09-139-675-2	Sequence 2, Appli
7	104.5	2.7	886	3 US-09-110-116-3	Sequence 3, Appli
8	103.5	2.7	1956	4 US-08-843-417-10	Sequence 10, Appli
9	103	2.7	438	2 US-08-677-049-9	Sequence 9, Appli
10	103	2.7	834	2 US-08-861-464-4	Sequence 9, Appli
11	103	2.7	834	2 US-09-336-001-4	Sequence 4, Appli
12	103	2.7	834	4 US-09-323-433A-4	Sequence 4, Appli
13	101	2.6	622	2 US-08-132-990A-4	Sequence 4, Appli
14	101	2.6	622	5 PCT-US92-09382-4	Sequence 4, Appli
15	99.5	2.6	462	2 US-08-898-976-2	Sequence 2, Appli
16	99.5	2.6	462	2 US-08-898-976-2	Sequence 2, Appli
17	99	2.6	574	1 US-08-898-976-4	Sequence 4, Appli
18	99	2.6	574	1 US-08-140-729A-7	Sequence 7, Appli
19	99	2.6	574	1 US-08-916-745-7	Sequence 7, Appli
20	99	2.6	574	2 US-08-948-569A-4	Sequence 4, Appli
21	99	2.6	574	2 US-08-663-808-6	Sequence 6, Appli
22	99	2.6	574	2 US-09-042-929-7	Sequence 7, Appli
23	99	2.6	574	2 US-08-546-661-7	Sequence 7, Appli
24	99	2.6	574	2 US-09-042-960-7	Sequence 7, Appli
25	99	2.6	574	2 US-09-188-469-4	Sequence 4, Appli
26	99	2.6	574	3 US-09-198-650-7	Sequence 7, Appli
27	99	2.6	574	3 US-09-332-740-6	Sequence 6, Appli

28	99	2.6	574	3 US-09-042-913-7	Sequence 7, Appli
29	99	2.6	574	3 US-09-188-496-6	Sequence 6, Appli
30	99	2.6	574	3 US-09-042-937-7	Sequence 7, Appli
31	99	2.6	574	4 US-09-397-238A-4	Sequence 6, Appli
32	96.5	2.5	401	3 US-08-492-459-6	Sequence 6, Appli
33	96.5	2.5	401	3 US-08-492-459-8	Sequence 8, Appli
34	96.5	2.5	401	3 US-08-492-459-28	Sequence 28, Appli
35	96.5	2.5	401	3 US-08-492-459-29	Sequence 29, Appli
36	96.5	2.5	401	3 US-08-423-752-6	Sequence 6, Appli
37	96.5	2.5	401	3 US-08-423-752-8	Sequence 8, Appli
38	96.5	2.5	401	4 US-08-716-873-6	Sequence 6, Appli
39	96.5	2.5	401	4 US-08-716-873-20	Sequence 20, Appli
40	96.5	2.5	401	4 US-08-716-873-22	Sequence 22, Appli
41	96.5	2.5	401	4 US-08-716-873-42	Sequence 42, Appli
42	96.5	2.5	401	4 US-08-716-873-43	Sequence 43, Appli
43	96.5	2.5	401	4 US-09-368-431-6	Sequence 6, Appli
44	96.5	2.5	401	4 US-09-368-431-20	Sequence 20, Appli
45	96.5	2.5	401	4 US-09-368-431-22	Sequence 22, Appli

## ALIGNMENTS

RESULT 1  
US-08-424-567-2  
; Sequence 2, Application US/08424567  
; Patent No. 5569755  
; GENERAL INFORMATION:  
; APPLICANT: SCHWEINFEST, Clifford W.  
; APPLICANT: PAPAS, Takis S.  
; TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated  
; TITLE OF INVENTION: Expression In Colon Adenomas And Adenocarcinomas  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424.567  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/026.045  
; FILING DATE: 05-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/181 NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 764 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-424-567-2

Query Match 28.9%; Score 1120; DB 1; Length 764;  
Best Local Similarity 34.0%; Pred. No. 1.7e-101;  
Matches 250; Conservative 142; Mismatches 275; Indels 68; Gaps 11;  
QY 7 RIVVDRAVSLTLFDDFEKKDRTIPVG-EKLRNAFRGSSAKIKAVFGLLPYLSWLMKRY 65

```

Db      8 QYIVARPVYSTNAFEENHKKTRGHHKTFDLHLKVCSCSPQAKAKIVLSTLFPJASLPAY 67
Qy      66 KIKDYIIPDLGLSGSGSIQVPOGMAFALLANLPAVNGLYSSFFPLTYFFFGVGHQWMP 125
Db      68 RKEWMLSDIYSGISTGIVAVLQGLAFALLVDIPVYGGLYASFPAIYLFEGTSHHSIV 127
Qy      126 GTFVAISILVY----NICQLAPESKFOVE---NNATNESYVDJAMEERLHVSATLAC 178
Db      128 GFFPLISMVWVGLAVGANSKAVPDRNATTLGLPNNSNSSLDD--ERVVAAAASYTV 184
Qy      179 LTAIOMGLGFMQGFVAVIYSESPFRGFMATAGLIILSVLYKIFGLTIPSTVGSGSIV 238
Db      185 LSGIILQALFGLIRIGFVVYIYSESLISGFTTAAAHVLSQLKFTFQLVPSHTDVSIF 244
Qy      239 FTFIDICKNLPHTNINSLIFALISGAFVLVYKELNARYMHKIRPPIPTEMIYVVAITAS 298
Db      245 KVLVSFQIEKTNIALDTALIVLIVVYSIVKEINQRFKDKLPVPIPIEFIMTVIAGVS 304
Qy      299 GGCKMPKRYHMOIYGEIQRGFPTPVSPVVSQKMDIGTAFSLAIVSYINLAMGRTLANK 358
Db      305 YGCDKRNKRYAVVDMDMNGFQPIPTPVETFTQNTVGDGFIAMVAFVAFVSASVYSIK 364
Qy      359 HGVDVDSNOEMIALGCSNPFSGFEKIHVICALSVTLAVDAGAGSKQVASCVSILVMT 418
Db      365 YDPLDGNDELIALGLNIVCGVFRGFAGSTALSRSAVOESTGKTQIALGLIGALIVLIV 424
Qy      419 MLVIGIYLYPLKPSVGLALIVANLKNLSKQLTDPYLYMRKSLDCCIIVWSFSLSSFLSL 478
Db      425 VLAIGFLAPLQKSVLALALGNLKMQLMGAIEIGRLMRKDYDCLIMTFIFITVIGL 484
Qy      479 PYGAVAVGAFSVLVVYFOTQFRNGYALAOVMDTIYVNPRTYNRAODIOGIKIITYCSPL 538
Db      485 GLGLAASVAFOLLITVFTQFPKCSSTLANIGRTINYNKKKDYDMEPEGVAKIFRCPSPI 544
Qy      539 YFANSEIFROKVI-----AKTVSLOELOO-----DPEANP 568
Db      545 YFANIGFRRKLLDVAVGSPLRLILKRNKALKIRKLOKQGLLOVTPKGFICTVDTIKDS 604
Qy      569 PTDPNNNO----TPANGTSVSY-ITFSPDSSSPAQSEBPASAEAPGEPSPDLMAVSPVP 622
Db      605 DEELDNQIEVLDPQINTTDLRPHIDMNDL-----PLNIEVPK-----I 644
Qy      623 TPHIILDMGSGSFVDLMGKIKALAKLSYIGKIGVAFVLVNHQVYNDISHGVEFEDGS 682
Db      645 SHSLILDFSAVSLDVSSVYRGLKSLIQEFIRIKVYIVGTDDFIEKINRYEEFF-DGE 703
Qy      683 LECKHVPSIHDAVL 697
Db      704 VKSSTIFLTHDAVL 718

RESULT 2
US-08-711-928-2
: Sequence 2, Application US/08711928
: Patent No. 5831015
: GENERAL INFORMATION:
: APPLICANT: SCHWEINERT, Clifford W.
: APPLICANT: PAPAS, Takis S.
: TITLE OF INVENTION: Colon Mucosa Gene Having Down-Regulated
: TITLE OF INVENTION: Expression In Colon Adenomas And Adenocarcinomas
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/711,928
: FILING DATE: 11-SEP-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/424,567
: FILING DATE: 17-APR-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/026,045
: FILING DATE: 05-MAR-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 40399/354/NIND
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 764 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-711-928-2

Query Match      28.9% Score 1120: DB 2: Length 764:
Best Local Similarity 34.0%: Pred No.1.7e-101:
Matches 250: Conservative 142: Mismatches 275: Indels 68: Gaps 11:

Qy      7 RYVDRRAVSLTLPFDEFEKDRTPVVG-EKLRNARFGSSAKIKAVFGLLPVLSWLPKY 65
Db      8 QYIVARPVYSTNAFEENHKKTRGHHKTFDLHLKVCSCSPQAKAKIVLSTLFPJASLPAY 67
Qy      66 KIKDYIIPDLGLSGSGSIQVPOGMAFALLANLPAVNGLYSSFFPLTYFFFGVGHQWMP 125
Db      68 RKEWMLSDIYSGISTGIVAVLQGLAFALLVDIPVYGGLYASFPAIYLFEGTSHHSIV 127
Qy      126 GTFVAISILVY----NICQLAPESKFOVE---NNATNESYVDJAMEERLHVSATLAC 178
Db      128 GFFPLISMVWVGLAVGANSKAVPDRNATTLGLPNNSNSSLDD--ERVVAAAASYTV 184
Qy      179 LTAIOMGLGFMQGFVAVIYSESPFRGFMATAGLIILSVLYKIFGLTIPSTVGSGSIV 238
Db      185 LSGIILQALFGLIRIGFVVYIYSESLISGFTTAAAHVLSQLKFTFQLVPSHTDVSIF 244
Qy      239 FTFIDICKNLPHTNINSLIFALISGAFVLVYKELNARYMHKIRPPIPTEMIYVVAITAS 298
Db      245 KVLVSFQIEKTNIALDTALIVLIVVYSIVKEINQRFKDKLPVPIPIEFIMTVIAGVS 304
Qy      299 GGCKMPKRYHMOIYGEIQRGFPTPVSPVVSQKMDIGTAFSLAIVSYINLAMGRTLANK 358
Db      305 YGCDKRNKRYAVVDMDMNGFQPIPTPVETFTQNTVGDGFIAMVAFVAFVSASVYSIK 364
Qy      359 HGVDVDSNOEMIALGCSNPFSGFEKIHVICALSVTLAVDAGAGSKQVASCVSILVMT 418
Db      365 YDPLDGNDELIALGLNIVCGVFRGFAGSTALSRSAVOESTGKTQIALGLIGALIVLIV 424
Qy      419 MLVIGIYLYPLKPSVGLALIVANLKNLSKQLTDPYLYMRKSLDCCIIVWSFSLSSFLSL 478
Db      425 VLAIGFLAPLQKSVLALALGNLKMQLMGAIEIGRLMRKDYDCLIMTFIFITVIGL 484
Qy      479 PYGAVAVGAFSVLVVYFOTQFRNGYALAOVMDTIYVNPRTYNRAODIOGIKIITYCSPL 538
Db      485 GLGLAASVAFOLLITVFTQFPKCSSTLANIGRTINYNKKKDYDMEPEGVAKIFRCPSPI 544
Qy      539 YFANSEIFROKVI-----AKTVSLOELOO-----DPEANP 568
Db      545 YFANIGFRRKLLDVAVGSPLRLILKRNKALKIRKLOKQGLLOVTPKGFICTVDTIKDS 604

```



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/02645  
FILING DATE: 01-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: STEEFE, ERIC K  
REGISTRATION NUMBER: 36,688  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-471-496-2

Query Match 3.2%; Score 123.5; DB 1; Length 470;  
Best Local Similarity 22.3%; Pred. No. 0.0015;  
Matches 105; Conservative 70; Mismatches 184; Indels 111; Gaps 21;

QY 35 EKLRAAFRCSSAKIKAVYGLL-----PVL-SWLPKRYKIKDYIIPDLGLSG 82  
DB 10 QRLKEGRASROLVAVYFVALLDNMLFTVVPVPTFLYMERKE-VISSLHGHAGS 68  
QY 83 SIQVPOGMFALLANLPAVNGLYSFFPLLYFELGCVHQMVPGEFAVISILVNICLOL 142  
DB 69 S---PHALA-----SPASTIF-SFNNNTV-----AVESVPSGLAMNDASTI----- 110  
QY 143 APESKFOVFNNATNSYVDTAAMEARLHVSATLCLAIITOM-----GLGF-- 189  
DB 111 -PPPATEAISAKNNCLOGTGFLBETTRVGVLFAS-KAVMOLVNPFGVPLTNRIGYHI 168  
QY 190 -MOGFVAITYLS-----ESFIRGFMTAGLOILISYL----- 220  
DB 169 PMFAGFVIMFLSTVWFARISGTYTLTFVARTLOGISSFSVAGLMLASVYDDHERGRA 228  
QY 221 --KYIFGLTIPSYTGP--GSIVFTFIDICKNLPHNTIASLIFALISGAFVLVKELNARY 276  
DB 229 MGTALGGLALGLGVAPGVSVMYEVG--KSAPFLILAF--ALLDGAQLCLOLQPSKYS 284  
QY 277 MHKIR----FPIPTEMIVVAVTAISGGCKMPKTHMOIVGEIQRGFPPTVSPVVSQMKD 332  
DB 285 PESAGTFLFMLKDPYILVAAGSI---C-----FANMGVAILEPTLPIMMQTCSPRW 336  
QY 333 MGTAFSLAIVSYVINLAMGRFLANKHGVDVDSNDEMIALGCSNFGSFFKTHVICALS 392  
DB 337 QGLGLFLEPASVSYLIGTNLFGVLANKMGRLCSLIGMLVGTG-----LLC 382  
QY 393 VTLAVDAG--GKSQVASLCSLVVMITMLVGIYLPKSVLGAALAV 440  
DB 383 VPLAKHKNGLIGPNAGLGLXIGMVESMMPIHGHLVDRPHTSVYGSVAHI 432

RESULT 5  
US-08-894-840-2  
Sequence 2, Application US/08894840  
Patent No. 5859200  
GENERAL INFORMATION:  
APPLICANT: LI, YI  
APPLICANT: CAO, LIANG  
APPLICANT: ROSEN, CRAIG A.  
TITLE OF INVENTION: HUMAN AMINE TRANSPORTER  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 29-AUG-1997  
APPLICATION NUMBER: US/08/894,840  
ATTORNEY/AGENT INFORMATION:  
NAME: STEEFE, ERIC K  
REGISTRATION NUMBER: 36,688  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-894-840-2

Query Match 3.2%; Score 123.5; DB 2; Length 470;  
Best Local Similarity 22.3%; Pred. No. 0.0015;  
Matches 105; Conservative 70; Mismatches 184; Indels 111; Gaps 21;

QY 35 EKLRAAFRCSSAKIKAVYGLL-----PVL-SWLPKRYKIKDYIIPDLGLSG 82  
DB 10 QRLKEGRASROLVAVYFVALLDNMLFTVVPVPTFLYMERKE-VISSLHGHAGS 68  
QY 83 SIQVPOGMFALLANLPAVNGLYSFFPLLYFELGCVHQMVPGEFAVISILVNICLOL 142  
DB 69 S---PHALA-----SPASTIF-SFNNNTV-----AVESVPSGLAMNDASTI----- 110  
QY 143 APESKFOVFNNATNSYVDTAAMEARLHVSATLCLAIITOM-----GLGF-- 189  
DB 111 -PPPATEAISAKNNCLOGTGFLBETTRVGVLFAS-KAVMOLVNPFGVPLTNRIGYHI 168  
QY 190 -MOGFVAITYLS-----ESFIRGFMTAGLOILISYL----- 220  
DB 169 PMFAGFVIMFLSTVWFARISGTYTLTFVARTLOGISSFSVAGLMLASVYDDHERGRA 228  
QY 221 --KYIFGLTIPSYTGP--GSIVFTFIDICKNLPHNTIASLIFALISGAFVLVKELNARY 276  
DB 229 MGTALGGLALGLGVAPGVSVMYEVG--KSAPFLILAF--ALLDGAQLCLOLQPSKYS 284  
QY 277 MHKIR----FPIPTEMIVVAVTAISGGCKMPKTHMOIVGEIQRGFPPTVSPVVSQMKD 332  
DB 285 PESAGTFLFMLKDPYILVAAGSI---C-----FANMGVAILEPTLPIMMQTCSPRW 336  
QY 333 MGTAFSLAIVSYVINLAMGRFLANKHGVDVDSNDEMIALGCSNFGSFFKTHVICALS 392  
DB 337 QGLGLFLEPASVSYLIGTNLFGVLANKMGRLCSLIGMLVGTG-----LLC 382  
QY 393 VTLAVDAG--GKSQVASLCSLVVMITMLVGIYLPKSVLGAALAV 440  
DB 383 VPLAKHKNGLIGPNAGLGLXIGMVESMMPIHGHLVDRPHTSVYGSVAHI 432

RESULT 6  
US-09-139-675-2  
Sequence 2, Application US/09139675A  
Patent No. 6117426  
GENERAL INFORMATION:  
APPLICANT: LI, YI  
APPLICANT: CAO, LIANG  
APPLICANT: ROSEN, CRAIG  
TITLE OF INVENTION: Human Amine Transporter  
FILE REFERENCE: 1488 0830003  
CURRENT APPLICATION NUMBER: US/09/139,675A  
CURRENT FILING DATE: 1998-08-25

```

; EARLIER APPLICATION NUMBER: WO PCT/US95/02645
; EARLIER FILING DATE: 1995-03-01
; EARLIER APPLICATION NUMBER: US 08/471,496
; EARLIER FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (402)
; OTHER INFORMATION: May be any amino acid
US-09-139-675-2

```

```

Query Match          3.2%; Score 123.5; DB 3; Length 470;
Best Local Similarity 22.3%; Pred. No. 0.0015;
Matches 105; Conservative 70; Mismatches 184; Indels 111; Gaps 21;

```

```

QY 35 EKLNAFRCSSAKIKAVFGLL-----PVL-SMLPKYIKDYIIPDLGLSGG 82
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 10 QRLKEGASQOLVLYVFFVALLDNMLFTYVVPYIYPTFLDMERKE-VISSLHGHS 68
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 83 SIYPOGNAFALLANLPAVNGLYSSFFPLTYFFLGVHGVNPGTEFAVISILVNICQL 142
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 69 S---PHALA-----SPAFSTIF-SFNNNTV---AVEESVPSGIAMNDTASTI--- 110
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 143 APESKFOVNATNESYVDYDAMEERLHVATLACLTATIIOM-----GLGF-- 189
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 111 -PPPATEAISAHKNCLOGTGFLEETTRGVLEAS-KAVNQLVNPVGPPLTRNIGYHI 168
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 190 -MDFGFVAIYLS-----ESFIRGFMTAGLOIILSVL----- 220
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 169 PMFAGFVIMETSTWFAFSGTYTLFVARTLOGIGSSFSVAAGLMLASVITDDHERGRA 228
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 221 --KIYFGLTIDSYGP--GSIVTFPIDICKNLPHNTIASLIFALISGAEVLVKELNARY 276
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 229 MGTALGIALGLLVGAPGVSVMYEVG--KSAPFLILAFL--ALLDGAQLQCIIDPSKYS 284
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 277 MHKIR----PIPTEMIVVVVATATASGCGMKPKRYHMOIYGEIOMGFPTPVSPPVSOMKD 332
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 285 PESAKGTPLEMLDPPYLLVAGSI--C-----FANNGVAILLEPTLIDMMQOTCSPKW 336
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 333 MICTAFSLAISYVYINLMGRTLANKHGYDVDSNOEMIALCSCNFGSFFKIHVICALS 392
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 337 QLGIAFLPASVSYLIGTLNLEGVLANKGRWLCSLIGMLVGTG-----LLC 382
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 393 VTLAVDAG--GKSQVASLQVSLVVMITMLVGLIYLPKPSVLCALIAV 440
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 383 VPLAHKNGLGPNAGLGLXIGWESSMMPIMGHLVDPRHRSYVGSVAHI 432
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

```

RESULT 7
US-09-110-116-3
; Sequence 3, Application US/09110116
; Patent No. 6013479
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong
; APPLICANT: Cohan, Victoria L.
; APPLICANT: Stuart, Susan G.
; TITLE OF INVENTION: HUMAN EMK1-LIKE G PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR
; FILE REFERENCE: PF-0550 US
; CURRENT APPLICATION NUMBER: US/09/110.116
; CURRENT FILING DATE: 1998-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 866
; TYPE: PRT
; ORGANISM: HOMO SAPIENS

```

```

; FEATURE:
; OTHER INFORMATION: 784994, GenBank
US-09-110-116-3

```

```

Query Match          2.7%; Score 104.5; DB 3; Length 866;
Best Local Similarity 20.7%; Pred. No. 0.32;
Matches 63; Conservative 51; Mismatches 105; Indels 85; Gaps 14;

```

```

QY 264 AFLVYVKELNARYMHKIFPIPTEMIVVVVATATASGCGMKPKRYHMOIYGEIOMGFPTPV 323
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 482 SFVGMESVLENERFDDHQAPLTTSIKLAKNSRYVGGI-----MTGEKKDGFSDPI 532
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 324 -----SPVVSQMK-DMIG--TAFSLAIY---SYVI-----NLAM--- 351
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 533 IYTLLENVQPKQKFERPICVSMSTDVKGGRMTSFGCVILLEASPTYIICSCNOMANLAVIMA 592
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 352 -GRTLANKHGYDVDSNOEMIALG-----SNFF-----GSFFKIIV-ICCALSVTLA 396
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 593 SGEITMDSLYIISHVGIISLVLAIAFLLCRSIRNMHTYLAHLVCVCLLAKTLF 652
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 397 VDAGAGKSOVASLQVSLVVMITMLVGLIYLPKPSVLCALIAVNLKSLQOTDPPYIYM 456
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 653 LAGI-HKTDNKTGCAIINGFLHYLFLACFPMLVAVILFLMVRNK----- 698
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 457 RRSKLDCIIVVVSFLSFFSLPYGVAVGVAFSVLVVVFQFR-NGYALAQVMDFTIYV 515
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 699 -----VYNYRSSHNIKMLHICAFGYGLPMLVVISASVQPGVG---MHNRCWL 744
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QY 516 NPKT 519
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
DB 745 NTET 748
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

```

RESULT 8
US-08-843-417-10
; Sequence 10, Application US/08843417
; Patent No. 6184349
; GENERAL INFORMATION:
; APPLICANT: Herman, Ronald C
; APPLICANT: Delgado, Stephen G
; APPLICANT: Fish, Linda M
; APPLICANT: Sangameswaran, Lakshmi
; APPLICANT: Rabert, Douglas K
; TITLE OF INVENTION: CLONED PERIPHERAL NERVE
; TITLE OF INVENTION: TETRODOXIN-RESISTANT SODIUM CHANNEL alpha-SUBUNIT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehnman White & Mcauliffe
; STREET: 525 University Ave
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,417
; FILING DATE: April 15, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Schmonsees, William
; REGISTRATION NUMBER: 31,796
; REFERENCE/DOCKET NUMBER: 28340-P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)-324-7041
; TELEFAX: (415)-324-0638
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1956 amino acids

```

TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-843-417-10

Query Match 2.7%; Score 103.5; DB 4; Length 1956;  
 Best Local Similarity 19.4%; Pred. No. 1.5;  
 Matches 112; Conservative 93; Mismatches 244; Indels 127; Gaps 22;

QY 122 QNPGTEFVITLVGNICQLAPESKFOVENNATNESYVDTAAMEARLHSATLA--CL 179  
 DB 576 ELAPGAVDVSAFDAGCKTFLSAEYLDPEFRAQRAMSVSITTSVLEELSECKCPCL 635  
 QY 180 TAIIO-----MGLGQRFQFVAIYLSSEFIRGFMTAGIQLISVLKTYF---- 224  
 DB 636 TSLSKYLIWDCPCPMVKLTFLGLVDPFAELTI-----TLCIVNTTFMAME 685  
 QY 225 --GL--TIPSYTPGSIYFTFIDICKNLPHTNIALIFALISGAFVLVKELNARYMKI 280  
 DB 686 HHGMSPTFEMALQIGNIYFTIF-----FTAEVFKIITA-----\*FDPIYYFQK 727  
 QY 281 RPIPTFEMIYVV---VATPISGCKMPKKYHMQIVGEIQRGFPPTPVSPVVSQMKDMIGT 336  
 DB 728 KWNIFDCIIVTVSLELGVAKKGSLSVLRSPFLRLRVFKLAKSMPT-LMTLKIKIGNSVGA 786  
 QY 337 ---AFSLAIYSYVINLAMPRTLANKHGYDVDSNOEMIAL-----GCSNFPSGFKRI 384  
 DB 787 LGNLTIIILATIVFVAL--VGKQL--GENYRNNRKNISAPHEDPVRMHMHDFHSFLIV 842  
 QY 385 HVICCALSVTLAVDAGGKQVASLCSLVVMTLVIGIYLYPLKSVLGLALIVNLKN 444  
 DB 843 FRILCG---EMENMMAMQMEVGKSIICLILEFTYVWIGNLV-----VLNFIALLNS 892  
 QY 445 -SLKQLTDPYILMRKSKIDCCIWVSF-----LSSEF--LSLPIYGAVGAVSFLVVV 494  
 DB 893 FSADMLTLPEDDGEVNNILQVALRIQVFGHRTKQALCSFFSRSCFPPOPKAPELVVXLP 952  
 QY 495 FOTFRNGYALQVM-----DIDIYVNPRTYNNRQDLOGIKIILYYCSPL 538  
 DB 953 LSSSAENHIANTRKSGSGGLQAPRGPRDEHSDFIANFTYVWSPVPIEGESDLDLEDD 1012  
 QY 539 YFANSEIFRQKVIAT--TVSLQELQODENAPPTDPNNNQTPANGTSVYITFSPDSSSP 596  
 DB 1013 GGEPAOSRQOEVIYPKGOEOQLQVVERCGDHLTP-----RSPGCTSSDDLAPSLGETWK 1066  
 QY 597 AQSEPPASAEAPG-----EPSDMIASVP 619  
 DB 1067 DESVQAPAEQVDTSSEGSTVDCLDPEELRKIP 1102

RESULT 9  
 US-08-677-049-9

Sequence 9, Application US/08677049  
 Patent No. 5858707  
 GENERAL INFORMATION:  
 APPLICANT: Guimaraes, M. Jorge  
 APPLICANT: Bazan, J. Fernando  
 APPLICANT: Mcclanahan, Terrell K.  
 APPLICANT: Zlotnick, Albert  
 TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;  
 TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DNA Research Institute  
 STREET: 901 California Avenue  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94304-1104

COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/677,049  
 FILING DATE: 03-JUL-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/000,788  
 FILING DATE: 03-JUL-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ching, Edwin P.  
 REGISTRATION NUMBER: 34,090  
 REFERENCE/DOCKET NUMBER: DX0511  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-852-9196  
 TELEFAX: 415-496-1200  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 438 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: Region  
 LOCATION: 115..144  
 OTHER INFORMATION: /note= "Encompasses TM 4 of Figure 4"  
 FEATURE:  
 NAME/KEY: Region  
 LOCATION: 323..357  
 OTHER INFORMATION: /note= "Encompasses TM 9 of Figure 4"  
 FEATURE:  
 NAME/KEY: Region  
 LOCATION: 359..386  
 OTHER INFORMATION: /note= "Encompasses TM 10 of Figure 4"  
 OTHER INFORMATION: 4"  
 US-08-677-049-9

Query Match 2.7%; Score 103; DB 2; Length 438;  
 Best Local Similarity 19.6%; Pred. No. 0.14;  
 Matches 80; Conservative 76; Mismatches 132; Indels 120; Gaps 22;

QY 172 VSATLACITAIQMGLOGMQGFVVAIYLSSEFIRGFMTAGIQLISVLKTYIFGLTIPSY 231  
 DB 74 LGCTFTAVSPMAIG--SEYGVSTVY-----GSIIASG--ILVILISFEFKLV-SF 120  
 QY 232 TGP---GSIV---FTFIDICKNLPHTNIA-----SLIPALISGAFV----- 266  
 DB 121 PFPVVTGSVVTITGITLMPVAMN---NMAGEGADGDSLNLAATVTSIIVLYRF 176  
 QY 267 -----VLKELNARYMKIRPIPTFEMIYVVVAVATPISGCKMPKKYHMQIVG 313  
 DB 177 TKGFIKSVSILIGLIGFIYFMCKVQFDNVSDAAV-----QMIO 218  
 QY 314 ELQKGFPT-PVSPVVSQMKDMIGTAFSLAIVSY----INLAMPRTLANKHGYDVDSQOE 368  
 DB 219 PFFGAPSFHAAPIT-----TMSIYAIYSLVESTGVYFALG-DLNNRRLTEIDLSKG 269  
 QY 369 MIALGCSNFGSPFKIHVICCALSVTLAVDAGGKQVASLCSLVVMTLVIGIY-LY 427  
 DB 270 YRABGLAVLGGTIN-----AFPTT-AFSQNVGLVQLTGTKKNAVIVYTVIIMARGLF 322  
 QY 428 P-----LPKSVLG---ALIAVNLKNSLKQLTDPYILMRKSKL--DCCIVV---VSF 470  
 DB 323 PKIAFTTIIPISAVLGGAMVAMFGVIAVYGIKMLSRIDFAQENLLIYACSVGLGVTY 382  
 QY 471 LSSFFLSLP-----YGAVVAVSFLVVVFTOTFRNGYALQVMD 510



ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-6408A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 834 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-001-4

Query Match 2.7%; Score 103; DB 2; Length 834;  
Best Local Similarity 17.6%; Pred No. 0.41; Indels 250; Gaps 33;  
Matches 140; Conservative 110; Mismatches 294;

QY 93 ALLANLPAVNGLYSFFPLTYFPL-GGVHQMVPQTFAVI-----SILVGNICL 140  
DB 87 ASMAMPAL-----LPLINEFDLEMDGPRKSSHDFTVAPNSGVNTSSLIMETPS 139  
QY 141 QLAPEKQOVF---NNATNESYVDTAAMEAEERLHVSATLACTAITIOMGLGM-----OF 192  
DB 140 SVTPAASLRNFNSNNMAKSCGVNDSFGLSS-STSSSWEISALPLRLDYIKLATDQF 198  
QY 193 G--FAIVL---SES-FIRGFMTAGLOLISLVKYGELTIPSTGPGSIVFTFIDICK 246  
DB 199 GCRFLQKLETPESESNNVADLMYEQIKPFPLDLIDPFG-----NLYOKLCD 246  
QY 247 NLPHTNIALFALISGAFVLVYKELNARYMKIRFPITPTMIVVAVATAISGGCKMPK 306  
DB 247 YLTAEQKTLTIOTIYPNVFOISINOYGRSLQKIIDTVONEVDILLI----- 294  
QY 307 YHMQIVGELQGRFP---TPVSPVVSQMKDMCTAFLAIVSYVINLAKMRTLANKHGIDV 363  
DB 295 -----KGFSEFTSIEQVTVLINDLNG-----NHVIOKCIJFKSPSKFGFLI 336  
QY 364 DS--NOEMIALGCSNPFSGFFKIHVICALSVTLAVDAGAGSOVASLCVSLVVMITMLV 421  
DB 337 DAIVQNNITITISTHKG-----CCYLQKLSVCTL---QOIFKISVKIVQFLPGLI 385  
QY 422 ---LGIIVLYPLPKSV-----LGAIIAVNLKNSLKQLDPPYLYMKRSKLDCCIVWVSFLS 472  
DB 386 NDQFNNTIIOFLDIKELDEYLLAEFLNRLSNELCOLS-----CLKRSSNVV 432  
QY 473 SFELSLPYGAVGV-----ANSV-LVVVFQTOF 499  
DB 433 EKFIKRLRIITGFIYNNNGASQRTAASDDVINASMNILLTTIDIFTVINLVLIRDF 492  
QY 500 RRGVYLAQVMDTDIYVNPRTYNRADIDG----- 528  
DB 493 GN-YALQTLDDVKNSPLLAIVKNSNAIGONSSSTLNGNFCNDSLSKIGNILVLTKELL 551  
QY 529 -----IKTIYCSPLYEPAENSEIFROKVIYAKTVSLQELQODFEN----- 566  
DB 552 PSIKTTYSARKIKELVKVAYAEATGIPETDISPOVYAMSHNNQITINNEKNKPHNNSHNH 611  
QY 567 -----APPTDPNNNQ-----TPANG---TSVSYIT--FSPDSSSPAQSEPPA-- 603  
DB 612 NHHNNHNNHNNNNNNKSHTRHPSLRANAYHRSNSSVTNFNSNOYAOODKIKHSPDI 671  
QY 604 -----SAEAPGEPDMLASVPVTFPHLLIDMSGVSPVDLMGKALAKLSSTGKIGV 657  
DB 672 MNEFONAIYPSKCAPSFNSQTNPLVVS-HNSLQNFNDROFANLMAHPNSAPLHSSSSNI 730  
QY 658 KVLVNIHAQYNDISHSQVF--EDGSLECKHVFPSIDAVLFAQANARDVTPG----- 709  
DB 721 ---TNVAPNYSRGFGKQGFMMNETDKINAHNFSF-----YSNANQNFNESEVPRMQ 779  
QY 710 HNFQAGAPDAELSL 723

DB 780 YQTEGANMDSLSM 793

RESULT 12  
US-09-323-433A-4  
Sequence 4, Application US/09323433A  
Patent No. 6218512  
GENERAL INFORMATION:  
APPLICANT: Guarante, Leonard P.  
APPLICANT: Austriaco Jr., Nicanor  
APPLICANT: Claus, James J.  
APPLICANT: Cole, Francesca  
APPLICANT: Kennedy, Brian  
TITLE OF INVENTION: GENES DETERMINING CELLULAR SENSENCE IN  
FILE REFERENCE: 0050, 1491-003  
CURRENT APPLICATION NUMBER: US/09/323,433A  
CURRENT FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: US 08/396,001  
PRIOR FILING DATE: 1995-02-28  
PRIOR APPLICATION NUMBER: PCT/US94/09351  
PRIOR FILING DATE: 1994-08-15  
PRIOR APPLICATION NUMBER: US 08/107,408  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 834  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-09-323-433A-4

Query Match 2.7%; Score 103; DB 4; Length 834;  
Best Local Similarity 17.6%; Pred No. 0.41; Indels 250; Gaps 33;  
Matches 140; Conservative 110; Mismatches 294;

QY 93 ALLANLPAVNGLYSFFPLTYFPL-GGVHQMVPQTFAVI-----SILVGNICL 140  
DB 87 ASMAMPAL-----LPLINEFDLEMDGPRKSSHDFTVAPNSGVNTSSLIMETPS 139  
QY 141 QLAPEKQOVF---NNATNESYVDTAAMEAEERLHVSATLACTAITIOMGLGM-----OF 192  
DB 140 SVTPAASLRNFNSNNMAKSCGVNDSFGLSS-STSSSWEISALPLRLDYIKLATDQF 198  
QY 193 G--FAIVL---SES-FIRGFMTAGLOLISLVKYGELTIPSTGPGSIVFTFIDICK 246  
DB 199 GCRFLQKLETPESESNNVADLMYEQIKPFPLDLIDPFG-----NLYOKLCD 246  
QY 247 NLPHTNIALFALISGAFVLVYKELNARYMKIRFPITPTMIVVAVATAISGGCKMPK 306  
DB 247 YLTAEQKTLTIOTIYPNVFOISINOYGRSLQKIIDTVONEVDILLI----- 294  
QY 307 YHMQIVGELQGRFP---TPVSPVVSQMKDMCTAFLAIVSYVINLAKMRTLANKHGIDV 363  
DB 295 -----KGFSEFTSIEQVTVLINDLNG-----NHVIOKCIJFKSPSKFGFLI 336  
QY 364 DS--NOEMIALGCSNPFSGFFKIHVICALSVTLAVDAGAGSOVASLCVSLVVMITMLV 421  
DB 337 DAIVQNNITITISTHKG-----CCYLQKLSVCTL---QOIFKISVKIVQFLPGLI 385  
QY 422 ---LGIIVLYPLPKSV-----LGAIIAVNLKNSLKQLDPPYLYMKRSKLDCCIVWVSFLS 472  
DB 386 NDQFNNTIIOFLDIKELDEYLLAEFLNRLSNELCOLS-----CLKRSSNVV 432  
QY 473 SFELSLPYGAVGV-----ANSV-LVVVFQTOF 499  
DB 433 EKFIKRLRIITGFIYNNNGASQRTAASDDVINASMNILLTTIDIFTVINLVLIRDF 492  
QY 500 RRGVYLAQVMDTDIYVNPRTYNRADIDG----- 528  
DB 493 GN-YALQTLDDVKNSPLLAIVKNSNAIGONSSSTLNGNFCNDSLSKIGNILVLTKELL 551





TELEPHONE: 202 628-5197  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 622 amino acids  
 TYPE: AMINO ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US92-09382-4

Query Match 2.6%; Score 101; DB 5; Length 622;  
 Best Local Similarity 20.5%; Pred. No. 0.4;  
 Matches 96; Conservative 60; Mismatches 206; Indels 106; Gaps 17;

QY 97 NLPVAVGLSSFFLLTYFLGVS-----HQMVPCTFAVISILY-----GNT 138  
 DB 155 NMPGVLATQPDIFAVIIIIITGLTLGKESAMVKIFTCINVLVLCFLIVSGEYKSTI 214  
 QY 139 CLQLAPESKFOVFNNATNESYVDTAMEAEERLHV-SATLACLTAITOMGLGFMOGFEVAT 197  
 DB 215 KNNQLEKMFSCNNDNTNKKYGGGFMPPFGSGVLSGANTCFYAFV-----GFDCT 265  
 QY 198 YLSESEFIRGFMPTAGLIQILISLVKYLFGLTIPSYTGPSTIV-----FTFIDICKNP--- 249  
 DB 266 ATGGEVKNPKOKAIPVGIASLL-----ICFIAYFGVSALTLMPYFCIDIDISPLPGAF 320  
 QY 250 -HTNINSLIFALISGAFVLVVELNARVMHKIRFPIPTKMIYVVAFTAISG-----GCKK 303  
 DB 321 KIQGWEAKYVAAGISGLALSTSLGS-----MFPMPR-----VIYMAEDGLLKFELAKI 371  
 QY 304 PKYHQIYGEIQRGFPPTVSPVVSQMKD-----IGTAFSLAIVSYI-----N 348  
 DB 372 NNRTKTPVATATYSGAIAVMAFLFELKDLVDLMSIGTLLASLVAACVLVRYPQQR 431  
 QY 349 LAMGRTLANKHGYDVDSNOEMIA-----LCCSNFEGSGFVKI 384  
 DB 432 LVMQARTTEELDRVDONELVSASESQTGFLPYAEKFSLSKLSIPKRVPSKFSGLIVNI 491  
 QY 385 HV-ICGALSVTLAVDAGKSOVASLCSLY-----VMITMVLGILYLYPLPKSVLGL 437  
 DB 492 SAGLALALITVTCIYAVLRELAEGTLMAVFMVMTGSVLLCMVLTII-TWROESKTKLS 550  
 QY 438 IAVNLKNSLKQLDPPYLMRKSRLDCCINW---VSFLSFFLSLPGV 482  
 DB 551 FKPVPVPLVLSIFVNIYLMQLDGTWVRFAVMMLIGFTIYFGIGI 598

RESULT 15  
 US-08-898-976-2  
 Sequence 2, Application US/08898976  
 Patent No. 5891670  
 GENERAL INFORMATION:  
 APPLICANT: Burnham, Martin  
 APPLICANT: Lonetto, Michael  
 APPLICANT: Warren, Patrick  
 TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE  
 TITLE OF INVENTION: PROTEIN  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Dechert Price & Rhoads  
 STREET: 997 Lenox Drive, Building 3, Suite 210  
 CITY: Lawrenceville  
 STATE: NJ  
 COUNTRY: USA  
 ZIP: 08543  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/898,976

FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bloom, Allen  
 REGISTRATION NUMBER: 29,135  
 REFERENCE/DOCKET NUMBER: GM10044  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 609-520-3214  
 TELEFAX: 609-520-3259  
 TELEX:

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 462 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

US-08-898-976-2

Query Match 2.6%; Score 99.5; DB 2; Length 462;  
 Best Local Similarity 17.7%; Pred. No. 0.34;  
 Matches 83; Conservative 72; Mismatches 121; Indels 193; Gaps 19;

QY 110 PLTYFELGGMVQMGCTFAVISILVGNICQLAPESKFOVFNNATNESYVDTAMEAEER 169  
 DB 25 PILLFLVFSL-VINSRKLSVAL-----ADD 51  
 QY 170 LHVASATL-----ACTTAIIOMGLGFMOGFEV-AIYLSSEFIRG----- 207  
 DB 52 LINSVTVSMQATLAV-IGMGAVVYASLSDAISIRPPFIYGVLLIRFSIIGNIFQH 110  
 QY 208 -----MTAGLIQILISLVKYL-----FGLTIPST-----TGCGSIV 238  
 DB 111 SPLTLVGRITOTAGLAAETLVYIVYAKYLSKEDQTYLGLSTSYSLVIGTISGFI 170  
 QY 239 FPFID-----ICKNLP-----HTNINSLIF-----ALISGAFVLVREL 272  
 DB 171 FTYLHMTNMFALIVYFTLPFLKLLPKENNTNKAHLDFVGLLVAITATYVMTLFTNF 230  
 QY 273 N-----ARYMHKIRFPIPT-----MIVVVAATAISG-----C 301  
 DB 231 NMLYMGALIALIVFALYIKNAQRPVKNKSFQNKRYASFLLFVGVYAIQLGYIFTFPE 290  
 QY 302 KMPKRYHMQI-----VGEIQGFPTPPVSPVVSQMKDMGTAF 338  
 DB 291 IMEQIYHQLDPTSLLLVPGYIVAVIAGALSGKIGEYLNKQAIITAIILALSLILPAF 350  
 QY 339 SLA-IVSYVINL-----AMGRTLANKHGYDVDSNOEMIALGCSNFGSGFVKIHYIC 388  
 DB 351 AVGNHISIFVISIMFFAGSFALMYAPLNALEAITDLNMTGAIGYVNL-----IIN 402  
 QY 389 CALSVTLAVDGA-----GKKSQVASLCSLVMTMVLVGLIY 427  
 DB 403 VAVSVGIAIAALIDRKALNFGNDALSSHFGIILILIGMSIVGLVF 451

Search completed: April 26, 2002, 09:05:37  
 Job time: 203 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 09:01:49 ; Search time 28.07 Seconds  
(without alignments)  
1987.075 Million cell updates/sec

Title: US-09-749-589-2  
Perfect score: 3869  
Sequence: 1 MSQPRPRVVDRAAYSLTFL.....MDLEQEMFGSMFHAETLAL 753

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1259.5	32.6	758	21	AAV71067 Human membrane tra
2	1116	28.8	764	15	AA860568 Down-regulated in
3	974.5	25.2	739	20	AAV08928 Rat DTDSF protein.
4	972	25.1	739	20	AAV08928 Human DTDSF protein.
5	735.5	19.0	506	22	AAAM2772 Peptide #1706 enco
6	735.5	19.0	506	22	AAAM2773 Peptide #1774 enco
7	735.5	19.0	506	22	AAAM3017 Peptide #1699 enco
8	601.5	15.5	143	22	AAAM2394 Human polypeptide
9	576	14.9	593	21	AAV44945 wheat sulphate per
10	566	14.6	680	21	AAV44943 Soybean sulphate p
11	566	14.6	688	21	AAV44935 Corn sulphate perm

12	549	14.2	631	21	AA629247
13	543.5	14.0	658	21	AA631621
14	543.5	14.0	658	21	AA632200
15	543.5	14.0	703	21	AA631620
16	538.5	13.9	621	21	AAV44942
17	537	13.9	656	21	AAV44944
18	498	12.9	579	21	AAV44936
19	452.5	11.7	842	21	AAV44938
20	437	11.3	746	21	AAV44358
21	425	11.0	565	22	AAAB76817
22	425	11.0	579	22	AA690917
23	403.5	10.4	606	22	AAE04904
24	385.5	10.0	499	21	AA629248
25	355.5	9.2	510	21	AA631622
26	355.5	9.2	510	21	AA632201
27	332.5	8.6	446	21	AA629249
28	288	7.4	457	21	AA632202
29	255	6.6	483	22	AA683019
30	253	6.5	537	22	AA690717
31	253	6.5	537	22	AA676818
32	240	6.2	76	21	AA603678
33	237.5	6.1	128	21	AA640320
34	225	5.8	466	21	AAV95728
35	220.5	5.7	311	21	AAV44937
36	201	5.2	390	19	AAW98528
37	191	4.9	411	22	AA682493
38	179.5	4.6	485	22	AA691372
39	179.5	4.6	485	22	AA676816
40	172.5	4.5	312	21	AA642187
41	170	4.4	582	21	AA607557
42	163	4.2	55	22	AAAM1978
43	163	4.2	55	22	AAAM3220
44	161	4.2	416	20	AAV37263
45	151.5	3.9	226	21	AAV44938

#### ALIGNMENTS

RESULT 1	AAV71067	Human membrane transport protein, MTRP-12.
ID	AAV71067	standard; Protein: 758 AA.
XX	AAV71067	
AC	AAV71067	
XX	XX	
DT	29-AUG-2000	(first entry)
XX	XX	
DE	Human membrane transport protein, MTRP-12.	
XX	XX	
KW	Human; membrane transport protein; MTRP-12; antiinflammatory; cytosolic; antithyroid; immunosuppressive; thymimetic; antidiabetic; nootropic; antidiarrhetic; neuroprotective; antidepressant; nephrotoxic; virucide; antihelminthic; protozoacide; antibacterial; neuroleptic; antiout;	Arabidopsis thalia Arabidopsis thalia Arabidopsis thalia Arabidopsis thalia Soybean sulphate p Wheat sulphate per Corn sulphate perm P. chrysogenum sut Corynebacterium gl C glutamicum prote Human transporter Arabidopsis thalia Arabidopsis thalia Arabidopsis thalia Arabidopsis thalia S. epidermidis ope C. glutamicum ope Corynebacterium gl Human secreted pro Human OREF ORF84 p Cosmid CHR15 enco Corn sulphate perm H. pylori GPO 141 S. epidermidis ope C glutamicum prote Corynebacterium gl Human OREF ORF951 Protein encoded by Peptide #6232 enco Peptide #7257 enco Chlamydia trachoma Artichoke sulphate
KW	diagnosis; prevention; treatment; membrane transport disorder; epilepsy; Alzheimer's disease; diabetes; Parkinson's disease; neurological disorder; inflammatory disorder; AIDS; Addison's disease; atherosclerosis; gout; Graves disease; Hashimoto's thyroiditis; microbial infection; cancer; cell proliferative disorder.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
PH	Key	Location/Qualifiers
FT	Modified-site	138
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	174
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	251
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	376
FT	Modified-site	/note= "phosphorylation site"
FT	Modified-site	574
FT	Modified-site	/note= "phosphorylation site"



XX	Key	Location/Qualifiers
FT	Misc-difference	620..640
FT	Misc-difference	/note="transcription activation domain"
FT	Misc-difference	566..573
FT	Misc-difference	/note="potential nuclear targeting motif"
FT	Misc-difference	573..580
FT	Misc-difference	/note="potential nuclear targeting motif"
FT	Misc-difference	576..583
FT	Misc-difference	/note="potential nuclear targeting motif"
FT	Region	460..764
FT	Region	/note="COOH terminal region"
FT	Region	1..175
FT	Region	/note="NH2 terminal with 4 Asp-linked glycosylation sites"
XX	W09420616-A.	
XX	PD	15-SEP-1994.
XX	PE	04-MAR-1994; .94WO-US01860.
XX	PR	05-MAR-1993; 93US-0026045.
XX	PA	(SCHW/.) SCHWEINFEST C W.
XX	PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	PI	Papas TS, Schweinfest CW;
XX	DR	WPI; 1994-303023/37.
XX	DR	N-PSDB; AAQ71399.
XX	PT	Isolated polypeptide that is down regulated in colon
XX	PT	adenocarcinomas and adenomas - is used as an indicator of tissue
XX	PT	abnormality.
XX	PS	Claim 3; Page 25-27; 41pp; English.
XX	CC	DRA can be used as antigens for the production of monoclonal
XX	CC	antibodies, which may then be used diagnostically for detecting
XX	CC	the presence/absence of DRA polypeptide in tissue sample. Absence
XX	CC	of DRA protein indicates tissue abnormality.
XX	Sequence	764 AA;

```

Query Match          28.8%: Score 1116: DB 15: Length 764:
Best Local Similarity 33.9%: Pred. NO.7.9e-104:
Matches 249: Conservative 142: Mismatches 275: Indels 68: Gaps 11.

Qy 7 RYVVDRAVSYLTLEPDEEKKDRRTYPVG-EKLRNAFRCCSAKIKAVVGLLVLVSLWPKY 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 qyIaVrPyvSYnafeenkkctgrhkctfdhkvccscspqakrtivlslfjiaawvlpay 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 66 KIKDYIIPDLGLGSGSIOVPOGMAFALLANPAVNGLSFFPPLYEFLGVHQWVP 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 rIkewllsIdivsgIsctgiVavIqafafalIdaIrpvygyasfIpaIlyIfgtsrshIev 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 126 GTFANISILVQ---NICLOLAPESKPYQF---NNATNSYVDYTMAMEERLHVSAATLAC 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 gPflIImmvglavsgavskavpIdnatlglIpmsnmsIldd--eTvraaaasvvc 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 179 LTAIIOMGLGFMQGFVAIYISESFIRGEMTAGLIILSVYKIFYGLIPSYTGGGSIV 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 lsgIIqlaIfglIrIgrIvIlyIseslsIsgftaaavhIvIsqkIfIqlIvpehtDpvsIf 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 239 FTFDIDCKNLPRPTNIASLFALISGAFVLVKEVLNARKYMKIRFPITPTMIYVVAATAIS 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 kvIysvfgIekIctnIdaIvtalIvIlyIsaIveIdqrIdckIrpvIptIefIntvIaaGys 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 299 GGCCKPKRYHMOIYGEIORGFPPTPVSPVYSOKMDMGTFASLAISYVNLAMGRTLAKK 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 ygcIdkntfkIkaavvampgpfIqplIpdvctIqnIvgdcfGIamaIafavaIsvaSyIsIk 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

**OY** HGYVDNSNOEIALTGCSNFESPEKIHVICALSVTLADVGAGSKSQVASTCYSLVMIT 418  
:: : ||| :: |:: : ||| :| :|||:  
**Dd** 365 Ydrpdqgqellalgnlignivcgvirfgagstalsrsavgescgkctqaglgaillvly 424  
: : : ||| ||| : : ||| : ||| :  
**Oy** 419 MLVGVITYLPSPKSVLALLAVNLKNLSLKQLTDPYIYMRSKLDDCCIMWVSFLSFEELS 478  
: : | : ||| ||| : : ||| : ||| : ||| :  
**Dd** 425 Vlaiqlfllapdqkvlaalaalnlgmllmqfaelgrlrwrkdycdlilwmtflftlvgl 484  
: : : ||| : : ||| : ||| : ||| :  
**Oy** 479 PYGAIVGVAESEVLVVYRQTQFRNCRGTALAQMWDIDVIYNPTTYNRADIQGKITTCSPSL 538  
: : ||| : : ||| : ||| : ||| : ||| :  
**Dd** 485 glglaasvaeflllelvrtfqfpkcstlanigrtnilynkdkdydmypgepvklfrpspi 544  
: : ||| : : ||| : ||| : ||| :  
**Oy** 539 YEANSEIRROKVI-----AKTVSLQEELAQO-----DENAP 568  
||||| :||| :  
**Dd** 545 yfanagfirrrikldavgfsplrlirknkalrkrkrgkgllqvrvpkgticvcldtkds 604  
: : : ||| : : ||| : ||| : ||| :  
**Oy** 569 PTDPNNNQ----TPANGTSVSX-ITFSPDSSSPAQSEPPASAEAPCEPSDMLASVPPEV 622  
605 deeldnqgievlidgpintltidpfhndwnddl-----plnievpk-----l 644  
: : ||| : : ||| : ||| : ||| : ||| :  
**Oy** 623 TFHTILDMSGSVSYEDLMGIFALKALSSTYGKTICVKPFLVNHAQVYNDISHGVFEDGS 682  
: : ||| ||| ||| : : ||| : ||| : ||| :  
**Dd** 645 slshllildsfavsflcdsvsgvjksllqeifririkvdviyvgtdddfleklnryeiff-dge 703  
: : ||| : : ||| : ||| : ||| : ||| :  
**Oy** 683 LECKHVFPISINDAV 696  
: : : ||| ||| :  
**Dd** 704 vkssifitindav 717  
: : : ||| ||| :  
**RESULT 3**  
**AAY08929**  
**ID AAY08929 standard; Protein; 739 AA.**  
**AC AAY08929;**  
**XX 19-AUG-1999 (first entry)**  
**DT XX Rat DTDSR protein.**  
**DE XX**  
**XX DTGST; human. expression vector; sulphate transporter; screening;  
KM bone disease; cartilage disease; treatment; cell sulphate ion intake;  
KW drug preparation; rat.  
XX  
OS Rattus sp.  
XX JP11146790-A.  
PN PD 02-JUN-1999.  
XX PF 18-NOV-1997; 97JP-0335157.  
XX PR 18-NOV-1997; 97JP-0335157.  
PA (SUMU ) SUMITOMO SEIYAKU KK.  
XX MPI; 1999-378999/32.  
DR N-PSDB; AAX78074.  
XX Sulfate transporter gene expression vector  
PX Claim 2; Page 12-13; 22pp; Japanese.  
PS This invention describes the construction of a novel vector for sulfate  
transporter expression containing a DNA sequence encoding a mammalian  
sulfate transporter (expression product of the pDPSR gene) and  
containing no DNA sequence of 3' translation region of the mammalian  
sulfate transporter gene. The invention also describes: (A) a method for  
screening a human bone/cartilage disease treating agent including the  
steps: (1) transforming an animal cell with the above vector, (2)  
culturing the animal cell in the presence of a sample and (3) detecting  
the increase in the sulfate ion intake to the cell; (B) a drug  
preparation for the treatment of human bone/cartilage disease containing**

CC the above vector as the active component. The sulfate transporter  
 CC gene-containing vector is high in expression efficiency. This sequence  
 CC represents the rat DTOST protein used in the method of the invention.  
 XX  
 XX  
 SQ Sequence 739 AA:

Query Match 25.2%; Score 974.5; DB 20; Length 739;  
 Best Local Similarity 31.7%; Pred. No. 1.7e-89;  
 Matches 219; Conservative 140; Mismatches 280; Indels 51; Gaps 8;

OY 33 VGEKLRNFRCSAKIKAVFGLLPVLSMLPKYKIKDYIIPDLGLSGSIQVPGMAF 92  
 DB 70 VMKIKQSCQCAKIKRNIIFQIFPLTWLPKPYDLKKNIGMMSGLLVGLIIVPQSLEY 129  
 OY 93 ALLANLPAVNGLYSSFFPLTFEFLGQHWMPGFPAVSIIVGN-----CQL-A 143  
 DB 130 SLLAGQEPYGLYTSFAISLYIFGSRHSVGLFGICLMLGEVDLHKACPDLOT 189  
 OY 144 PESKFOVENN--ATNESYVDTPAMEAEERLHVSATLACTAIOMGLGFMQGFVAIYLS 200  
 DB 190 TASSLAFNGCVVNVNHLIDGLCKSCYALKIGSTVFLNAGYVYAMGFFGYGFSVYLS 249  
 OY 201 ESFINGFMTAAGLOLLISVLKIRFGLTIPSTYSGSYVTFPDICKNLPHTNIALSLIFAL 260  
 DB 250 DALLSGFYGAFFLLISGAKYLLGLSPRNGVSVLTWILHFRNHNKHLNCLDILTSI 309  
 OY 261 ISGATLVYKLELNARVMHRIPIPEMIVVAVATASGCGMPKKYHQIAGEIORGP 320  
 DB 310 LCLLYVPTKEINLEYFKSLIPAPITELLVVAATLASHFGKINENYISAGQIPCTGM 369  
 OY 321 TPVSPVSVQMKDMIGTAFSLAIVSYVINLAMGRITLANKHGYDVDSNOEMIALGCSNFPS 380  
 DB 370 PPQAPDWSLIPNVADALATSLIGFALIVSLEMFAKKHGYLVKANGEMAYGIFCNLPS 429  
 OY 381 PFKIHVICALSVTLAVDAGCKRSOVASLCVSLVMITMLVIGIYLPKSVLCALLAV 440  
 DB 430 FHCITSAALAKLIVKESTGCTGLSALVTSVILLVILLIAPLFGYSIQKCVGLIVLIV 489  
 OY 441 NLKNSLKQLTDPYLVMRKSKLDCCIMVVSFLSPFLDPYGAAGVAFSVLVVVOGTOR 500  
 DB 490 NLRGALLKFRDIPKMWISRMGLVLFVLMSSALLSTELIGLVYCVLSMFCVLLITQMP 549  
 OY 501 NGYALQVMDTDIVYNPKTYNRADIDGKITTYCSPFLFANSEIFROKIVAKTYSLOEL 560  
 DB 550 KSLIIGLEESSELESFESIKYKLRSGKJGKFRFIAPLYLNKECKSALKYKLLNPVLV 609  
 OY 561 QODEFENAPPTDNNNOTFANGTSVSYITFSPDSSPAQSEPPASAEAPCEPDMLASVPP 620  
 DB 610 KAAMKKAARKLKEET-----VLFHGD-----PDEVSMQISHDP- 643  
 OY 621 FVTFHTLIDMSGVSFVDLGIKALAKLSTSTGKIGVKKVFLVNIHAQVNDISHGCVFD 680  
 DB 644 -LELHTVLDCAIGFIDTAGIHLTKEVRYDGAIGIYLLIAGCNPVSVDLSAKGEY--- 699  
 OY 661 GSLECKH---VPPSHDAVLFAQANARD 705  
 DB 700 ---CKEENELLFYSLSAFAESQKE 725

RESULT 4  
 AA08928  
 ID AA08928 standard; protein: 739 AA.  
 XX  
 AC AA08928;  
 XX  
 XX 19-AUG-1999 (first entry)  
 XX  
 DE Human DTOST protein.  
 KW DTOST: human, expression vector; sulphate transporter; screening;  
 KW bone disease; cartilage disease; treatment; cell sulphate ion intake;  
 KW drug preparation.

XX Homo sapiens.  
 OS JP1146790-A.  
 PN 02-JUN-1999.  
 PD 18-NOV-1997; 97JP-0335157.  
 PF 18-NOV-1997; 97JP-0335157.  
 PR (SOMU) SUMITOMO SEIYAKU KK.  
 PA WPI, 1999-378999/32.  
 PS Sulfate transporter gene expression vector  
 PS Claim 2; Page 10-11; 22pp; Japanese.  
 XX  
 CC This invention describes the construction of a novel vector for sulfate  
 CC transporter expression containing a DNA sequence encoding a mammalian  
 CC sulfate transporter (expression product of the DTOST gene) and  
 CC containing no DNA sequence of 5' translation region of the mammalian  
 CC sulfate transporter gene. The invention also describes: (A) a method for  
 CC screening a human bone/cartilage disease treating agent including the  
 CC steps: (1) transforming an animal cell with the above vector, (2)  
 CC culturing the animal cell in the presence of a sample and (3) detecting  
 CC the increase in the sulfate ion intake to the cell; (B) a drug  
 CC preparation for the treatment of human bone/cartilage disease containing  
 CC the above vector as the active component. The sulfate transporter  
 CC gene-containing vector is high in expression efficiency. This sequence  
 CC represents the human DTOST protein used in the method of the invention.  
 XX  
 XX  
 SQ Sequence 739 AA:

Query Match 25.1%; Score 972; DB 20; Length 739;  
 Best Local Similarity 31.2%; Pred. No. 3e-89;  
 Matches 226; Conservative 138; Mismatches 285; Indels 76; Gaps 11;

OY 3 QRPKRYVDRAVSLTLPDEFKKDRTP--VGEKLRNFRCSAKIKAVFGLLPVLS 60  
 DB 49 RPYHRIILERQ-----EKEDLNKEFVKKLQKNCQSPAKAKMILGIFLPVIG 97  
 OY 61 WPKYKIKDYIIPDLGLSGSIQVPGMAFALLANLPAVNGLYSSFFPLTFYFLGCV 120  
 DB 98 WLPKYDLKKNIGDWMGSLIYGLIIVPQSLEYSLIAGEPYGLYTSFAISLYIFLGTS 157  
 OY 121 HQMVPQTEFAVISILVGNIC--LQ-----LABESKFOYFNNTNESYVDTAAM 165  
 DB 158 RHISVGLIFGICLMLIGETVRELQKAGYDHAHSAPSLIGVNSGSLIHLTSQRLCKSCY 217  
 OY 166 EAEERLHVSATLACTLAIITOMGLGFMQGFVAIYLSSEFIRGTMATAGLOLLISVLKIRF 225  
 DB 218 ---ALMVGSVLFVAGYVYAMGFFGYGFSVYLSDALISGFTVGSFLLISGAKYLLIG 274  
 OY 226 LTRPSYSGSYVTFPDICKNLPHTNIALSLFALISGAFVLVLELNARVMHRIPIRP 285  
 DB 275 LNPRTNGYSLITCWINHFRNHNKHLNCLDILTSILCLVILLIPTKELNEHFKSKIPAP 334  
 OY 286 TEMIVVAVATAISGCGMPKKYHMOIVGEIORGFPTPVSPVVSQMK---DMIGTAFSLAI 342  
 DB 335 IELVVVAATLASHFGKILHENYSLAGHPIPTGF---MPKQVEMNLIIPSVADVATAISL 391  
 OY 343 VSYVINLAMGRITLANKHGYDVDSNOEMIALGCSNFPSFFKIHVICALSVTLAVDAGAG 402  
 DB 392 IGFALIVSLEMFAKKHGYTVKANGEMAYGIFCNLIIPSFHFTCSAALAKLIVSESTQC 451  
 OY 403 KSOVASLGVSLVMITMLVGLIYLPKPSYGVGALIAVLAKNSLKQLTDPYLVMRKSKID 462  
 DB 452 HTGISGVNTAVILLIIVLIAPLFGYSIQKSVGLIVLIVNLRGALKKFRDLPKMSISLMD 511  
 OY 463 CCIMVVSFLSPFLPYGAVAGVAFSVLVVVOGTORNGVALQVMDTDIVYNPKTYNR 522

**D**b           || | ||| || | : ||| || : : ||         | : :::: : |  
512     twvwtlmhsallsteglllyvcfcifcylilrtqkpkssllglveesvfsvayxn 571  
  
**Oy**           553 AODIGIKIITFYGSPLYFANSEIFROKVIATKYVSLQELQDFENAPPTDPNNNQFANGT 582  
          572 lqtgpgikliffvaplylnkeckfsalysqvtnpllikwamkaakrkikekvcllgtl 631  
  
**Oy**           563 SVSYTFPSSSRQSPPASACABCPBPSDMLASVPFFVTFTLLIDMSGVFLYLMGI 642  
          632 q-----demsqvlshdp-----lehltilvidcsaiqlfldtagl 664  
  
**Oy**           643 KALKLSTGYCKICVKVFVLNIHQVVNDSHGVSFEEDSLECKH-----VFSIHDAVL 697  
          665 htkevrrdyeaiaigyllaqcnptvrslngey-----ckkeenllfysvyema 717  
  
**Db**           698 FAOAN 702  
          718 faevs 722  
  
**RESULT**       5  
**ID**           AAM15272  
          AAM15272 standard; Protein: 506 AA.  
  
**XX**           AAM15272;  
  
**DT**           12-OCT-2001 (first entry)  
  
**DE**           Peptide #1706 encoded by probe for measuring cervical gene expression.  
  
**XX**           Probe: human: microarray; gene expression; cervical epithelial cell;  
**KM**           cervical cancer.  
  
**XX**           Homo sapiens.  
**OS**  
**XX**           WO200157278-A2.  
**RN**  
**PX**  
**PF**          09-AUG-2001:  
**DD**  
**PR**          30-JAN-2001; 2001WO-US00670.  
**XX**  
**PR**          04-FEB-2000; 2000US-0180312.  
**PR**          26-MAY-2000; 2000US-0207456.  
**PR**          30-JUN-2000; 2000US-0608408.  
**PR**          03-AUG-2000; 2000US-0632366.  
**PR**          21-SEP-2000; 2000US-0234687.  
**PR**          27-SEP-2000; 2000US-0236359.  
**PR**          04-OCT-2000; 2000GB-0024263.  
**PA**  
**XX**           (MOLE-) MOLECULAR DYNAMICS INC.  
**PI**  
**DR**          Penn SG, Hanzel DK, Chen W, Rank DR:  
**WP**: 2001-488901/53.  
**PT**          Human genome-derived single exon nucleic acid probes useful for  
**PT**          analyzing gene expression in human cervical epithelial cells -  
**PS**          Claim 27; SEQ ID NO 20098; 487pp; English.

Query Match	Similarity	Score	DB	Length
Best Local Match	164	31.2%	Pred No. 1.6e-65	506
Matches	164	Conservative	105	Mismatches 212
			Indels	45
			Gaps	6
QY	185	MOIGMOPGFENVAIVYSESPFNGFPMFAAGLQILISVLYKFIKGLTIPSPGSIPTFIDI	244	
Db	1	vamgfifgvgfsvysldalgsfvgastflltsqakylglndprlmgvsgllctwlhv	60	
QY	245	CANLPHNTNIASLIFALISGAFVLVYKELNARKMKHRRFPIPEMIVVYATASGCKMP	304	
Db	61	fnnlhtknlcdltsllcllvpkckelnehkshkapidieivvvaatlashfgkjh	120	
QY	305	KYHMQIYGEIGRGPPTPSPVSWMK---DMIGTAFSAIYSYVYNLMDGRFLANKHGY	361	
Db	121	eynysiasghnptgtf---mmpkvppekmlipsavaadaiaislfgatvsiemfakkngy	177	
QY	362	DVDSNOEMIALGCSNFFSGFFKIHVICALLSVTLAVDGAQGSQVASLCSLVNMITMLV	421	
Db	178	lvkangemyaigfcmllpsffhncfttsaalaktllvkestgchtlqisgvvtalvlllvllv	237	
QY	422	LGIVLYPRKSVLGLHIAVNLKNSLKQLTDRPYLMRKSKLDCCIMVVSFLSSFSLPRG	481	
Db	238	lapiylsrlqsvlyglvtlvnlrgalrkrfdlprkmswisimdvlfvhlslsallsteig	297	
QY	482	VAVGAFAFSLVVPFOFNGFYALAQVMDTDIVNPKRTYNRAQDLOGIKITRYCSFLPYA	541	
Db	298	llvgvcfsifcyliltqrpkssllglveesefesvsayknlqtkpghkllffvapllyli	357	
QY	542	NSEIFRQKVIATVSLQELQDFENAPPTDPNNNOTPRANGTSVSYITFSPDSSSPAQSEP	601	
Db	358	nkecfksalqkqtnpjlilkvawkkakrklikekvcltggigq-----demsyqlshd	409	
QY	602	PASAAAPGSPSMLASVPFFVFPHLLIDMSGVSFVLDLMGIALAKLSSTYKIGYKVPFL	661	
Db	410	p-----lelhtlvldcsaigldtdaglnhtlkevtrdyeaigqlvll	450	
QY	662	VNIHAQVYNDISHGVEFEDGSLECKH---VFPSIHDAVLFPAQAN	702	
Db	451	aqcnpvtvrdslngey-----ckkeenallfyvsyemaataeys	489	
RESULT 6				
PA	AAM27737	standard; Protein; 506 AA.		
XX	AC	AAM27737;		
XX	DT	17-OCT-2001 (first entry)		
DE	XX	Peptide #1774 encoded by probe for measuring placental gene expression.		
XX	XX	Probe: microarray; human; placenta; antenatal diagnosis;		
XX	KW	genetic disorder.		
OS	XX	Homo sapiens.		
PN	XX	WO200157272-A2.		
XX	PD	09-AUG-2001.		
XX	XX	30-JAN-2001; 2001WO-US000663.		
PR	XX	04-FEB-2000; 2000US-0180312.		
PR	XX	26-MAY-2000; 2000US-0207456.		
PR	XX	30-JUN-2000; 2000US-0608408.		
PR	XX	03-AUG-2000; 2000US-0632366.		
PR	XX	21-SEP-2000; 2000US-0234687.		
PR	XX	27-SEP-2000; 2000US-0236359.		
PR	XX	04-OCT-2000; 2000GB-0024263.		
PA	XX	(MOLE-) MOLECULAR DYNAMICS INC.		

[illegible][illegible]



```

OY 542 NSRIFQKVIATKVSLOELQDFENAPPTDPNNQTPANCTSVYITFSDSSPAOSEP 601
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 358 nkecfksalykgtvnpillkvawkaakrkikxvltl99iq-----demsqishd 409
OY 602 PASAEAPGSPDMLASVPVFTFTLLDMSGVSPVDMGKRALAKUSTYKIGVVF 661
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 410 p-----lehtlvidsaigfidtagihlckerridyaaigvll 450
OY 662 VNIIAQVNDISHGVFEDEGLECKH-----VPSIHDAVLAQAN 702
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 451 aeqnptvrdsltngey-----ckkeenllfysvyeamafevs 489

RESULT 8
AA42394
ID AA42394 standard; Protein; 143 AA.
XX
AC AA42394;
XX
DE 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 127.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; antiviral;
KW antiparasitism; antischistosomal; antianemic; antitubercular; cancer;
KW antihelminthic; hepatotropic; cerebroprotective; antiinflammatory;
KW antidiabetic; antitubercular; antileishmaniasis; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234897.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
PI

```

```

XX WPI; 2001-476225/51.
DR N-PSDB; AA162799.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
XX Claim 11; SEQ ID NO 127; 532bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AA162752-AA162961) and proteins
CC (AA42347-AA42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 143 AA;

Query Match 15.5%; Score 601.5; DB 22; Length 143;
Best Local Similarity 91.0%; Pred. No. 8.9e-53;
Matches 122; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

OY 107 SFFPLTYFFLGGYHOMVPGTFAVISTLVGNICQLAPESKFOYFNATNESYDTAME 166
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 9 sffpltyffxgyvghnvpqtfavistlvgnicqlapeskfyfnatnesyvdaame 68

OY 167 AERLHVASATLACLTATITOMGFGFVAFIYSESRFRGFMATGQITISVYKTYFGL 226
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 69 aerlhwatlaclxlqmgiglmqrfvalysessflrgrmtaagqllslvkllyfga 128

OY 227 -TIPSYTGPSTIVE 239
    | : : : | | |
DB 129 gqepthswprslaf 142

RESULT 9
AA44945
ID AA44945 standard; Protein; 593 AA.
XX
AC AA44945;
XX
DE 23-MAY-2000 (first entry)
XX
DE Wheat sulphate permease-2.
XX
KW Sulphate permease; sulphate assimilation protein; wheat; probe;
KW mapping; marker; plant breeding; chimeric gene; transgenic plant;
KW antibody; screen.
XX
OS Trifolium aestivum.
XX
PN WO200004154-A2.
XX
PD 27-JAN-2000.
XX
PF 13-JUL-1999; 99WO-US15810.
XX
PR 14-JUL-1998; 98US-0092833.
XX

```





Db 601 -----khetlhfildsawpaidtsgiafild 628  
Qy 648 LSTYCKIGVKFVLNHAQVYNDISHGCVFED 680  
          : | | : | | : | | : | | :  
Db 629 lksstckrgtelvtnptgervmeklgraneen 661

RESULT 12  
AAG29247  
ID AAG29247 standard; Protein: 631 AA.  
XX AAG29247;  
AC  
XX  
XX 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34766.  
XX  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hydrolisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144336.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

```

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0159239.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159684.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 14.28; Score 549; DB 21; Length 631;

Best Local Similarity 26.08; Pred. No. 1.9e-46;

Matches 175; Conservative 120; Mismatches 251; Indels 126; Gaps 20;

```

QY 18 TLFDPEKKDRTPVGEKLNARFCSSAKIKAVFGLPVLWLPKIKIRYIIPDLG 77
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 ttfddpdrgrtgrgpnrkL-----lraagy-lfpilgwepeysf-sllksdvvs 72
QY 78 GLSGSIOVPOGMFALLANI.PAVNGLYSSFFPLLTFFLGGVHOMVGTFAVISILVGN 137

```

```

Db 73 gtlaslalpgalsyaklanlppivglvysltvplvavilssrdlavpvaslallgs 132
QY 138 ICIOLAPESKFOVFNNATNESYVDPAAMEAERLAHSATLCTAIITOMGLFMOPGFVAI 197
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 133 mlrq-----qv-----spvddpvflq-latsstf--faglfqslglrlrlgflld 175
QY 198 YLSESEFGEFMTAGLQILISVLYKIFGLITPSTYNGPSIVFTFDICKNLPHTN----- 252
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 176 flskatlglfmgaaaliavslgqlkgllgt--hftkmsvvpvlsav--lqhtneweqw 230
QY 253 --IASLIFALLSGAFVLVYKELNARYMKIRPPTTEMAIVVVAIAISGGCKMPKRYIMQ 310
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 231 flvmgvcflL---flslstrhsmkkpkrlfwvsagapilsvlsvcllvfvt-aerhgis 285
QY 311 IVGEIORGFPTPVSPVVSQMKD-----IGTAFSLAIVSVINLMDGRTLANKHG 360
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 vlgkipeglpp-----swmlglfghshlaivaktglvtglvstleglavgrftaaakn 339
QY 361 YDVSNOEMIALGCSNPFSGFEKIHVICCALSVTLAVDAGGKSOVASLVCVSLVVMITML 420
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 yhvdkmekmialglmrvvgsalscytlgafersavnnagaktavslmvsctvmvcll 399
QY 421 VLGIYLYPL---PKSVIGALIAVNLKNSLKOITDPYILMRKSKIDCCITWVSFLSFFL 476
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 ---flmplfeytpnvigall-vcaviglidlipachlwkdktfdlvmlocaifgvifl 454
QY 477 SLPYGAAGVAFSVLVVFOQFRNGYVALAQVMDIDIVNPFYRARDIOGIRITGCS 536
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 svnglalaavglstfkilmqvtrpkmvlmnglpgcdlrydlhnyeagrlpgflvlsles 514
QY 537 PLYFANSEIFROKVIKTVLSLOELOQDFENAPPTDPNNNGTPANGTSVSYTSPDSSSP 596
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 515 pvnfansnyltert---srwleecnee-----qflllemsavgydvtgvsffkelkktakkd 579
QY 557 AQSEPPASAEAPGERSDMLASVPPVTFRHTLLDMSGVSFVDMIKAKLAKSTYGGIG 656
   | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 539 -----eeqekhsjL-----qflllemsavgydvtgvsffkelkktakkd 579
QY 657 VKVFLVNIHAQV 668
Db 580 lelvtvmlpsev 591

```

RESULT 13  
AAG31621  
ID AAG31621 standard; Protein: 658 AA.

AC AAG31621;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38004.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141282.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147933.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149724.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 543.5; DB 21; Length 658;  
Best Local Similarity 24.9%; Pred. No. 7.5e-46;  
Matches 180; Conservative 127; Mismatches 274; Indels 141; Gaps 23;

OY 4 PPRPVVDRAAVSL--TLF-DDEFKKDRTPVGEKLRNAPFCSSAKIRAVFGL---LP 57  
DB 25 pppqpfllkslqysvkvellpddpfrq-----fkngnast-----kfvlglyklyflp 69  
OY 58 VLSWLPKTKIKYIIPDLGLSGSGSIQVPGCMARFALLANLPAVNGLYSFPPLTYFFL 117  
DB 70 lfwaprynlk-fkxslagltiaslaipglsyaklanlpllglyasfvpplvyavl 128  
OY 118 GGVHGVNPTFVVISILVGNICLQALAPESKQVFNNAATNESYVDATAMEAR----LHV 172  
DB 129 gssrdlavgltvvaalllg---amlske-----vdaekdkpkyllhl 166  
OY 173 SATLACLAIIOMGLGFMOGFVAIYLSSEFTRGFTAGLQILSLVKYIFGLTIPSYT 232  
DB 167 atfatffagvlaslglfflgfivdflshatlvgfmggaatlvsbqlglgfgl--khtf 224  
OY 233 GPGSIVFTFIDICKNLPHTNINSLIFALISGAFLVYLKELNARYMKIRPPTMTIVVY 292  
DB 225 dsltdvlsvmrsvfsqthewres---gvlscgflffl--lstry-----fsikpkkffwv 274  
OY 293 VATTA-----ISGCGCKM---PKYHMOIVGEIORGFPPTV-----SPVVSQMKDMIG 335  
DB 275 aamapltvslvsllyvtfhaerhgvqvgldlkgf--nplssdlfftsbyps---tavk 330  
OY 336 TAFSLAIVSYVINLAWGRTLANKHGYVDVSNQEMIALGCSNFGSPFKIHVTCALSVTL 395  
DB 331 tgljltgialaevagvgrsfamfkynldgnkemiafgmmnvgstfcylyltgpcrsra 390  
OY 396 AVDGAGGKQSVASLCVSLVVMITMLVGLIYLPKPSVUGALIAVNLKSLKOLTPPYL 455  
DB 391 vvnagcgktamenlymalavmtflflp1fhytplvlsail--lsamjldygaathl 449  
OY 456 WRKSKLDCCIMVVSFLSPFLSPYGAVAVFSVLVVVFOQFRNGVALAOVNDTDIVV 515  
DB 450 wkvdkfdflvcmsayygvvfgsvveiglvavalsiarlllfvsrpktavkgnpmismiy 509  
OY 516 NKRYNRAODIGIKITICYCSPLYPANSEIFRQKVIATAKTVSLOELOQDENAPPTDPNN 575  
DB 510 ntegypsrtvpgjllleldapylfanaasylerlirw-----ideeee 553  
OY 576 QTPRANC-TSVSVITTFPSDSSPRAQSEPPASAEAPGEPDOMLASVPPFVTFHTILDMSGV 634

DB 554 rvkqsgesslgy1-----lllmsav 573  
OY 635 SEVDLMGIKALAKLSSTYKIGKVFVLNIIHAQVYNDISHGVEDGSLECKHVPSPID 694  
DB 574 gnltdsgismmveikkvldrralkvlvsnpkgevvkkltrskffigd-hlgkewmfltyge 632  
OY 695 AV 696  
DB 633 av 634  
RESULT 14  
AAG32200  
ID AAG32200 standard; Protein: 658 AA.  
XX  
AC AAG32200;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38800.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0122180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0138492.  
PR 18-JUN-1999; 99US-0138434.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144083.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145228.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147410.

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 543.5; DB 21; Length 658;  
Best Local Similarity 24.9%; Pred. No. 7.5e-46;  
Matches 180; Conservative 127; Mismatches 274; Indels 141; Gaps 23;





```

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149179.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153736.
PR 13-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```

```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159768.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 14.0%; Score 543.5; DB 21; Length 703;  
 Best Local Similarity 24.9%; Pred. No. 8.3e-46;  
 Matches 180; Conservative 127; Mismatches 274; Indels 141; Gaps 23;

```

QY 4 PPRRYVDRAVSL--TLF-DDEFKKDRTTYVGEXLKRARFCSSAKTKAVVFG---LP 57
DB 70 PqPqPflKslgysvkellfpddpfr-----fkqgnasr-----kflgiklyfl 114
QY 58 VLSWLPKRYKIXDIIPDLGLSGSIOVPGMAFALLANLPVNGVLSFPLTYFEL 117
DB 115 lfwaprynlk-fkxdsllagtlstiaslapqdisyaklanlprlllgysfvpplvyavl 173
QY 118 GGVHQMVPOTFAVISILVGNICLOLAPESKFCQVFNNAATNESVVDRAAMEAER-----LHV 172
DB 174 gssrdlavgtvaeaslltg---amslke-----vdaekdpxlyhl 211
QY 173 SATLACLTAIIOMGICFMQFGFVAIYLSBSFIRGFWTAAGLOILISVUKYITGLTIPSYT 232
DB 212 aftarffagvleaslgifrlgflvdlshatlvgmgaactlvslqklrptgl--khft 269
QY 223 GPGSIVFFPIDICKNLPHTNINSLIFALISGAFVLVLELMARXYMKIRPRTPEMIVV 292
DB 270 dstdvismrsvfsqthetwex---gvlygcfllfl--lstry-----fsikpkpffwv 319
QY 293 VATA-----ISGCKM---PKYHMQIIVGEIORGFPPV-----SPVVSQMKDMIG 335
DB 320 aamapltvslgsllyvtfhaerhgqvlgldklgl-nplagsdillfsapyms---tavk 375
QY 336 TAFSLAIVSYVINLAMGRTLANKHGVDVDSNOEMTALGCSNFFGSEFFKIHVICALSVTL 395
DB 376 tglitgltalaegvavgsfamfknyldgnkemiafgmmnlvgfscycyltqpfgrsa 435
QY 396 AVDGAGKSQVASLCVSLVVMITMLVGLIYLPKPSVGLANLANLNKSLKOLIDPRYL 455
DB 436 vnyngacktamslvymalavmftllflcplrfhyrpllvlsali--lsamlgldygaahl 494

```

```

OY 456 WRKSKLDCCINWVSPFLSPFLSPGVANGVAFSVLVVVFOTQFRNGYALAOVMDTIDIV 515
DB 495 WKVdKfdIvcmSAYVGVVfgSvEIglvvaIstIrlIlfvsrpktaVkgnlpsmIyr 554
OY 516 NKRTYNRAODIOGKIITFYCSPLYFANSEIFROKVIKTVSLQELQDFENAPPTDPNNN 575
DB 555 nteqypsrtvpglllleIdapIyfanasyIlerIlrw-----Ideeee 598
OY 576 QTPANG-TSVSYITFSPDSSSPAQSEPPASAEAPGEPPSDMLASVPPFVTFHTLILDMSCV 634
DB 599 rvkgsgesIgyI-----lIdmsav 618
OY 635 SFVDLMGIRALAKLSSTYCKIGKVFVNIHAQVYNDISHGCVFEDGSLCKHVPPSIHD 694
DB 619 gnIdtsgIsmmveIkKvIdralkIvlnpkgvkvkItrskfIgd-hlgkewmfltyge 677
OY 695 AV 696
DB 678 av 679

```

Search completed: April 26, 2002, 09:05:16  
 Job time: 207 sec

